Taylor & Francis
Taylor & Francis Group

Check for updates

A Bayesian Variable Selection Approach Yields Improved Detection of Brain Activation From Complex-Valued fMRI

Cheng-Han Yu [a], Raquel Prado[a], Hernando Ombao[b], and Daniel Rowe[c]

[a]Department of Applied Mathematics & Statistics, University of California at Santa Cruz, Santa Cruz, CA; [b]Statistics Program, King Abdullah University of Science and Technology (KAUST), Saudi Arabia; [c]Department of Mathematics, Statistics and Computer Science, Marquette University, Milwaukee, WI

ABSTRACT

Voxel functional magnetic resonance imaging (fMRI) time courses are complex-valued signals giving rise to magnitude and phase data. Nevertheless, most studies use only the magnitude signals and thus discard half of the data that could potentially contain important information. Methods that make use of complex-valued fMRI (CV-fMRI) data have been shown to lead to superior power in detecting active voxels when compared to magnitude-only methods, particularly for small signal-to-noise ratios (SNRs). We present a new Bayesian variable selection approach for detecting brain activation at the voxel level from CV-fMRI data. We develop models with complex-valued spike-and-slab priors on the activation parameters that are able to combine the magnitude and phase information. We present a complex-valued EM variable selection algorithm that leads to fast detection at the voxel level in CV-fMRI slices and also consider full posterior inference via Markov chain Monte Carlo (MCMC). Model performance is illustrated through extensive simulation studies, including the analysis of physically based simulated CV-fMRI slices. Finally, we use the complex-valued Bayesian approach to detect active voxels in human CV-fMRI from a healthy individual who performed unilateral finger tapping in a designed experiment. The proposed approach leads to improved detection of activation in the expected motor-related brain regions and produces fewer false positive results than other methods for CV-fMRI. Supplementary materials for this article are available online.

1. Introduction

As an imaging modality, fMRI is able to indirectly measure neuronal activity by detecting changes in the blood oxygen level dependent (BOLD) signal. In a typical task-related fMRI experiment, hemodynamic activity over the entire brain volume is observed at T time points while a subject performs a series of tasks, leading to a set of T large-dimensional fMRI scans, typically T rectangular lattices with about 5K–10K voxels.
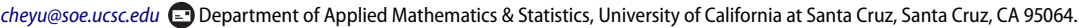
In MRI and fMRI, images or voxel measurements are complex-valued due to phase imperfections after Fourier encoding and inverse Fourier image reconstruction. Thus in fMRI, voxel time course measurements consist of real and imaginary components (Bernstein, Thomasson, and Perman 1989; Macovski 1996; Haacke et al. 1999) and these are generally converted to magnitude and phase voxel time courses. However, most fMRI brain activation studies discard the phase information and rely on magnitude-only image time courses. When this is done, the original complex-valued data are unrecoverable as operations that involve magnitude-only reconstruction are not unique. Some attempts have been made to avoid working with complex-valued voxel time courses or standard magnitude-based reconstruction algorithms. For instance, Bernstein, Thomasson, and Perman (1989) and Prah et al. (2010) showed that detectability in low signal-to-noise (SNR) regions of magnetic resonance images is improved by using a phase-corrected real reconstruction instead of magnitude-only reconstructions.

In this article, we develop a Bayesian model for detecting activation that uses both the real and imaginary components in CV-fMRI data, leading to more accurate activation results.

Bandettini et al. (1992) demonstrated that voxel time courses can be used as effective tools for localizing brain function in humans. Early common model-based approaches to the analysis of magnitude fMRI data relied on the general linear model (GLM), as first proposed by Friston, Jezzard, and Turner (1994). In this model, the observed magnitude-only fMRI signal is modeled as the underlying expected BOLD response plus a noise component. In other words, for each voxel $v = 1, \ldots, V$, the voxel-wise GLM can be written as

$$\mathbf{y}_v = \mathbf{X}_v \boldsymbol{\beta}_v + \boldsymbol{\epsilon}_v, \tag{1}$$

where $\mathbf{y}_v$ is the $T \times 1$ response vector of magnitude-only fMRI time course for voxel v, $\mathbf{X}_v$ is the $T \times q$ design matrix whose components include the expected BOLD responses for each of p experimental tasks or input stimuli and possibly other regressors such as trends (and so, $p \leq q$), $\boldsymbol{\beta}_v$ is a $q \times 1$ vector of regression coefficients and $\boldsymbol{\epsilon}_v$ is a $T \times 1$ error vector, which captures random noises due to scanner artifacts and any additional subject-related physiological noise. In the absence of intercepts, trends, or any other covariates that are not task-specific, that is, when $q = p$, each of the p BOLD responses in $\mathbf{X}_v$ is the discretized convolution of a stimulus on-and-off signal with the so-called hemodynamic response function (HRF) that models the

hemodynamic delay in the magnetic resonance signal (Friston et al. 2007). In addition, the HRF is often assumed to be the same across voxels, resulting in $\mathbf{X}_v = \mathbf{X}$ for all v.

Sophisticated Bayesian models, including spatial and spatio-temporal approaches, have been developed for magnitude-only fMRI data. For instance, Bowman et al. (2008) considered a two-stage Bayesian hierarchical model with temporal correlations at the first stage and spatial correlations at the second stage. In Lee et al. (2014), temporal dependence is characterized via autoregressive models, Zellner's g-priors are assumed for the regression coefficients, and a binary spatial Ising prior is used to specify anatomical information and spatial interaction between voxels. In Zhang et al. (2014), a general error structure is used to capture general dependence, and a Markov random field (MRF) prior is used to detect activations in a nonparametric way. Alternative Bayesian approaches for magnitude-only data are summarized in Zhang, Guindani, and Vannucci (2015), Zhang et al. (2016), and Chiang et al. (2017). These sophisticated and well-constructed models, however, are based only on the magnitude information provided by the data and do not incorporate the phase information. Furthermore, many of these magnitude-only approaches also work under the assumption that the errors are normally distributed which may be problematic, resulting in incorrect standard errors that can produce inaccurate activation results. In fact, if both the real and imaginary components of the CV-fMRI signals have independent normally distributed errors with the same variance, the magnitude-only signals actually follow a Ricean distribution that is approximately normal only in the case of large SNRs (Rice 1944; Gudbjartsson and Patz 1995; Rowe and Logan 2004). However, the SNRs may not be large enough in practice for this approximate normality to hold. This is increasingly true in cases with higher voxel resolutions and for voxels with a large degree of signal drop-out, that is, those for which the signal is not available or has small SNR, such as voxels located near air/tissue boundaries. In particular, Adrian, Maitra, and Rowe (2013) showed that with magnitude-only models, tests derived using Ricean modeling are superior to Gaussian-based activation tests for SNRs below 0.6. Rowe (2005b) also showed that Gaussian-based activation parameter estimates were biased for SNRs under 10. Our approach overcomes these limitations of magnitude-only models by jointly considering the real and imaginary components of CV-fMRI data.

Complex-valued modeling has been widely used in several applied areas allowing full utilization of real and imaginary, or equivalently magnitude and phase, information in certain signals and images, providing a general framework for the analysis of several classes of processes (see, e.g., Mandic and Goh, 2009). The incorporation of phase information has proven key in communications and imaging (Oppenheim and Lim 1981), as complex-valued modeling simultaneously handles the intensity and direction when dealing with radar, sonar, and wind data. In the fMRI context, CV-fMRI data that jointly consist of magnitude and phase images are not provided by the scanners as the default output, but they are usually readily available. For instance, GE scanners typically provide an output file that contains the raw complex-valued k-space data and other information, as well as the magnitude images. Magnitude and phase images, or real and imaginary images, can be easily

obtained by simply changing a preset control variable in an input file, making CV-fMRI data available to neuroimaging researchers and practitioners.

A number of tools for CV-fMRI data analysis have been proposed in the literature, including nonmodel-based exploratory independent component analysis (ICA; Calhoun et al. 2002), as well as direct modeling of the complex activation data (Lai and Glover 1997; Rowe and Logan 2004, 2005; Rowe 2005a; Lee et al. 2007; Rowe 2009; Lee, Shahram, and Pauly 2009). Approaches such as those in Rowe and Logan (2004, 2005); Rowe (2005a); and Rowe (2005b) model the phase to directly estimate the phase angle using a polar coordinates representation, while the methods in Lee et al. (2007) and Lee, Shahram, and Pauly (2009) are based on Cartesian representations. More recently, complex-valued models with temporal correlations (including autoregressive structures) have also been developed (Kociuba and Rowe 2016; Adrian, Maitra, and Rowe 2017). In particular, Rowe (2005a) specified the following structure for the complex-valued image measurement at time t and voxel v, $y_t^v = y_{t,Re}^v + i y_{t,Im}^v \in \mathbb{C}$,

$$y_t^v = \rho_t^v \cos\left(\phi_t^v\right) + i\rho_t^v \sin\left(\phi_t^v\right) + \eta_t^v, \qquad (2)$$

where $\rho_t^v = \beta_0^v + \beta_1^v x_{1,t} + \cdots + \beta_{p_1}^v x_{p_1,t}$ is the magnitude of y_t^v with p_1 magnitude regressors, $\phi_t^v = \alpha_0^v + \alpha_1^v u_{1,t} + \cdots + \alpha_{p_2}^v u_{p_2,t}$ is the phase of y_t^v with p_2 regressors, and $i = \sqrt{-1}$. All the regression coefficients $\beta_0^v, \ldots, \beta_{p_1}^v$ and $\alpha_0^v, \ldots, \alpha_{p_2}^v$ are real-valued. Here, a_{Re} and a_{Im} generically denote the real and imaginary parts of any complex-valued quantity $a = a_{Re} + ia_{Im}$. The noise term η_t^v is also assumed to be complex-valued, that is, $\eta_t^v = \eta_{t,Re}^v + i\eta_{t,Im}^v$. When $\alpha_0^v \neq 0$ and $\alpha_j^v = 0$ for all $j = 1, \ldots, p_2$, we have the Rowe-Logan constant phase model. Note that when no trends are included, the magnitude and phase regressors could be chosen to be identical to the expected bold responses associated with the p experimental tasks, that is, $p_1 = p_2 = p$ and $x_{j,t} = u_{j,t}$ for all $j = 1, \ldots, p$. Rowe (2005a) identified active voxels using a generalized likelihood ratio test.

Lee et al. (2007) and Lee, Shahram, and Pauly (2009) proposed a method based on a Cartesian model representation which has the following matrix form:

$$\mathbf{y}^v = \mathbf{X}\boldsymbol{\gamma}^v + \boldsymbol{\eta}^v, \qquad (3)$$

with $\mathbf{y}^v = (y_1^v, \ldots, y_T^v)'$, $\boldsymbol{\gamma}^v = \boldsymbol{\gamma}_{Re}^v + i\boldsymbol{\gamma}_{Im}^v$, $\boldsymbol{\gamma}_{Re}^v = (\gamma_{Re,1}^v, \ldots, \gamma_{Re,q}^v)'$, $\boldsymbol{\gamma}_{Im}^v = (\gamma_{Im,1}^v, \ldots, \gamma_{Im,q}^v)'$, with $q = p + 1$, $\mathbf{X} = (\mathbf{x}_1', \ldots, \mathbf{x}_T')'$, where $\mathbf{x}_t = (1, x_{1,t}, \ldots, x_{p,t})'$, $t = 1, \ldots, T$, and complex-valued noise vector $\boldsymbol{\eta}^v = (\eta_1^v, \ldots, \eta_T^v)$. Lee et al. (2007) combined this general linear model representation in Cartesian coordinates with a Hotelling's T^2-test to detect active sites. Model (3) is equivalent to the Rowe–Logan constant phase complex-valued model (Rowe and Logan 2004) if $p_1 = p$, $\boldsymbol{\gamma}_{Re}^v = (\beta_0^v, \ldots, \beta_p^v)' \cos(\alpha_0^v)$ and $\boldsymbol{\gamma}_{Im}^v = (\beta_0^v, \ldots, \beta_p^v)' \sin(\alpha_0^v)$. Model (3) is also equivalent to the complex-valued magnitude and phase activation model in Rowe and Logan (2005) when there is only a single regressor in both, magnitude and phase, corresponding to a 0/1 vector representing a boxcar block design.

The references cited above show that modeling the complete CV-fMRI data leads to superior power in detecting active voxels when compared to magnitude-only approaches, especially for

situations in which the SNRs are relatively small. However, in spite of their advantages, currently available methods for CV-fMRI data rely on mechanisms that control some notion of error to correct for multiple testing, such as Bonferroni corrections, and therefore involve two-step procedures. The first step provides estimates of the potentially active voxels according to some model, while the second step involves using one of the standard methods to correct for multiple testing. Furthermore, available methods for CV-fMRI data assume that the voxels are independent and do not offer a principled framework for parameter learning through borrowing information across voxels.

Here, we present a Bayesian approach that allows us to infer active voxels using both the real and imaginary information provided by the CV-fMRI data. This approach builds on Bayesian variable selection methods to detect active voxels and hence does not suffer from the multiple comparison issues that typically affect multiple hypothesis testing (Scott and Berger 2006). Activation detection and parameter estimation are achieved by a model-based framework that allows us to borrow information across voxels. In addition to obtaining full posterior inference via Markov chain Monte Carlo (MCMC), we develop a complex-valued extension of the Expectation-Maximization (EM) algorithm for Bayesian variable selection of Rockova and George (2014) that allows for fast detection of active voxels in large-dimensional CV-fMRI. The advantages of our approach are illustrated in the analysis of simulated data, including physically realistic simulated CV-fMRI data, as well as human CV-fMRI data. We show that the proposed methods lead to more accurate activation results than those obtained from magnitude-only methods or from currently available methods for CV-fMRI data. Section 2 presents the models and algorithms for posterior estimation and inference. Section 3 illustrates the performance of the Bayesian approach for detecting active voxels in simulated datasets, including physically realistic synthetic CV-fMRI data. Section 4 shows and discusses the results obtained from analyzing a human CV-fMRI dataset with the proposed Bayesian approach. Finally, Section 5 presents a discussion and future extensions.

2. Bayesian Models for Detecting Activation in Complex-Valued fMRI Data

As mentioned above, we develop a model that makes use of the complete magnitude and phase information provided by the CV-fMRI data. However, unlike previous approaches (Rowe and Logan 2004, 2005; Rowe 2005a, 2005b, 2009; Lee et al. 2007), we use a fully Bayesian framework for identifying active voxels via variable selection in the complex-valued domain.

We follow the Cartesian coordinates approach of Lee et al. (2007) given in (3) and further assume independent and identically distributed complex-normal error vectors, that is,

$$\mathbf{y}^v = \mathbf{X}\boldsymbol{\gamma}^v + \boldsymbol{\eta}^v, \quad \boldsymbol{\eta}^v \sim \mathrm{CN}_T\left(\mathbf{0}, \boldsymbol{\Gamma}_v, \mathbf{C}_v\right), \tag{4}$$

with $\mathrm{CN}_L(\boldsymbol{\mu}, \boldsymbol{\Gamma}, \mathbf{C})$ denoting a complex normal distribution of dimension L with mean $\boldsymbol{\mu}$, complex-valued, Hermitian and nonnegative definite covariance matrix $\boldsymbol{\Gamma}$, and complex-valued symmetric relation matrix $\mathbf{C}$. As shown below, the linear structure in this representation is computationally relevant, as it

leads to fast Bayesian posterior estimation of active sites. Note also that any complex-valued normal distribution of dimension L has a real-valued normal representation of dimension $2L$ (Wooding 1956; van den Bos 1995; Picinbono 1996). Thus, letting $\boldsymbol{\Sigma}_{Re,Re}^v = \frac{1}{2}Re(\boldsymbol{\Gamma}_v + \mathbf{C}_v)$, $\boldsymbol{\Sigma}_{Im,Im}^v = \frac{1}{2}Re(\boldsymbol{\Gamma}_v - \mathbf{C}_v)$, $\boldsymbol{\Sigma}_{Re,Im}^v = \frac{1}{2}Im(-\boldsymbol{\Gamma}_v + \mathbf{C}_v)$, and $\boldsymbol{\Sigma}_{Im,Re}^v = \frac{1}{2}Im(\boldsymbol{\Gamma}_v + \mathbf{C}_v)$, model (4) also has a real-valued representation as

$$\begin{pmatrix} \mathbf{y}_{Re}^v \\ \mathbf{y}_{Im}^v \end{pmatrix} = \begin{pmatrix} \mathbf{X} & \mathbf{0} \\ \mathbf{0} & \mathbf{X} \end{pmatrix} \begin{pmatrix} \boldsymbol{\gamma}_{Re}^v \\ \boldsymbol{\gamma}_{Im}^v \end{pmatrix} + \begin{pmatrix} \boldsymbol{\eta}_{Re}^v \\ \boldsymbol{\eta}_{Im,}^v \end{pmatrix}, \tag{5}$$

or equivalently,

$$\mathbf{y}_r^v = \mathbf{X}^r \boldsymbol{\gamma}_r^v + \boldsymbol{\eta}_r^v, \tag{6}$$

with $\mathbf{y}_r^v = ((\mathbf{y}_{Re}^v)', (\mathbf{y}_{Im}^v)')'$, $\mathbf{X}^r = \mathrm{blockdiag}(\mathbf{X}, \mathbf{X})$, $\boldsymbol{\gamma}_r^v = ((\boldsymbol{\gamma}_{Re}^v)', (\boldsymbol{\gamma}_{Im}^v)')'$, and $\boldsymbol{\eta}_r^v = ((\boldsymbol{\eta}_{Re}^v)', (\boldsymbol{\eta}_{Im}^v)')'$, where $\boldsymbol{\eta}_r^v \sim N_{2T}(\mathbf{0}, \boldsymbol{\Sigma}_v)$ with

$$\boldsymbol{\Sigma}_v = \begin{pmatrix} \boldsymbol{\Sigma}_{Re,Re}^v & \boldsymbol{\Sigma}_{Re,Im}^v \\ \boldsymbol{\Sigma}_{Im,Re}^v & \boldsymbol{\Sigma}_{Im,Im}^v \end{pmatrix}.$$

The simplest possible structure for $\boldsymbol{\eta}^v$ is that obtained by taking $\boldsymbol{\eta}^v \sim \mathrm{CN}_T(\mathbf{0}, 2\sigma_v^2 \mathbf{I}_T, \mathbf{0})$ or equivalently, setting $\boldsymbol{\Sigma}_v = \sigma_v^2 \mathbf{I}_{2T}$ in the real-valued Gaussian representation. This implies that there is no correlation within the real components and within the imaginary components of $\boldsymbol{\eta}^v$, and also that there is no correlation between the real and imaginary components of $\boldsymbol{\eta}^v$. These assumptions can be relaxed to include correlations within the real and imaginary components to capture temporal structure (as illustrated in some of the analysis of synthetic and human CV-fMRI data presented in Sections 3 and 4), or correlations between the real and imaginary components for more structured noise.

Below we describe the priors and the corresponding posterior inference for the simplest noise structure, focusing on complex-valued priors for $\boldsymbol{\gamma}^v$ that lead to posterior inference of activation in CV-fMRI at the voxel-specific level.

2.1. Priors

In the absence of any trends and intercepts, and without loss of generality, that is, for the case in which $\mathbf{X}$ in (4) contains only the expected BOLD signals for each of p stimuli/tasks with no baselines or trends, activation can be viewed as a variable selection problem (Xia, Liang, and Wang 2009; Zhang, Guindani, and Vannucci 2015). In other words, if $\gamma_j^v = \gamma_{Re,j}^v + i\gamma_{Im,j}^v \neq 0$ for voxel v and task j, such voxel is identified as active under task j. Note that complex-valued priors must be considered for γ_j^v. Here we develop a complex-valued domain analogue of the Bayesian variable selection methods of George and McCulloch (1993, 1997) and Rockova and George (2014). If trends and/or intercepts are needed, they can easily be included in the model along with priors on their corresponding parameters and integrated out, as done in the applications illustrated in Sections 3 and 4. Thus, we focus the discussion below to the case in which $\mathbf{X}$ only consists of the expected BOLD signals associated with each of the p experimental stimuli/tasks.

Our proposed complex-valued spike-and-slab priors for γ_j^v extend the widely used real-valued spike-and-slab priors

by considering

$$\left(\gamma_j^v \mid \psi_j^v\right) \sim \left(1 - \psi_j^v\right) g_0\left(\gamma_j^v\right) + \psi_j^v g\left(\gamma_j^v\right),$$

with $g_0(\cdot)$ and $g(\cdot)$ complex-valued distributions with mean zero, and $\psi_j^v \in \{0, 1\}$, where $\psi_j^v = 1$ indicates that voxel v is active during task j. Therefore, this prior allows us to determine if a voxel is active by jointly considering the real and imaginary components of γ_j^v. In general, we consider priors with $g_0(\gamma_j^v) = CN_1(0, \sigma_v^2 \omega_0, \sigma_v^2 \lambda_0)$, and $g(\gamma_j^v) = CN_1(0, \sigma_v^2 \omega_1, \sigma_v^2 \lambda_1)$, and their corresponding vectorial representation given by

$$\boldsymbol{\gamma}^v \mid \boldsymbol{\psi}^v \sim CN_p\left(\mathbf{0}, \sigma_v^2 \boldsymbol{\Omega}_v, \sigma_v^2 \boldsymbol{\Lambda}_v\right), \qquad (7)$$

with $\boldsymbol{\Omega}_v = \operatorname{diag}((1 - \psi_1^v)\omega_0 + \psi_1^v \omega_1, \ldots, (1 - \psi_p^v)\omega_0 + \psi_p^v \omega_1)$, $\boldsymbol{\Lambda}_v = \operatorname{diag}((1 - \psi_1^v)\lambda_0 + \psi_1^v \lambda_1, \ldots, (1 - \psi_p^v)\lambda_0 + \psi_p^v \lambda_1)$ and $\boldsymbol{\psi}^v = [\psi_1^v, \ldots, \psi_p^v]$. The real-valued representation of this prior is $\binom{\gamma_{Re}^v}{\gamma_{Im}^v} \sim N_{2p}(\mathbf{0}, \sigma_v^2 \boldsymbol{\Sigma}(\boldsymbol{\psi}^v))$, where $\boldsymbol{\Sigma}(\boldsymbol{\psi}^v) = \left(\begin{smallmatrix} \boldsymbol{\Sigma}_{Re,Re}(\boldsymbol{\psi}^v) & \boldsymbol{\Sigma}_{Re,Im}(\boldsymbol{\psi}^v) \\ \boldsymbol{\Sigma}_{Im,Re}(\boldsymbol{\psi}^v) & \boldsymbol{\Sigma}_{Im,Im}(\boldsymbol{\psi}^v) \end{smallmatrix}\right)$. Given $\boldsymbol{\psi}^v$, we obtain $\boldsymbol{\Sigma}(\boldsymbol{\psi}^v)$ from $\boldsymbol{\Omega}_v$ and $\boldsymbol{\Lambda}_v$ via $\boldsymbol{\Sigma}_{Re,Re}(\boldsymbol{\psi}^v) = \frac{1}{2}Re(\boldsymbol{\Omega}_v + \boldsymbol{\Lambda}_v)$, $\boldsymbol{\Sigma}_{Im,Im}(\boldsymbol{\psi}^v) = \frac{1}{2}Re(\boldsymbol{\Omega}_v - \boldsymbol{\Lambda}_v)$, $\boldsymbol{\Sigma}_{Re,Im}(\boldsymbol{\psi}^v) = \frac{1}{2}Im(-\boldsymbol{\Omega}_v + \boldsymbol{\Lambda}_v)$, and $\boldsymbol{\Sigma}_{Im,Re}(\boldsymbol{\psi}^v) = \frac{1}{2}Im(\boldsymbol{\Omega}_v + \boldsymbol{\Lambda}_v)$.

In the data analyses presented below, we take $g_0(\gamma_j^v) = CN_1(0, 2v_0\sigma_v^2, 0)$ and $g(\gamma_j^v) = CN_1(0, 2v_1\sigma_v^2, 0)$, with parameters $0 < v_0 < v_1$, and with smaller values of v_0 favoring the detection of even weakly activated voxels. As shown in Section 2.2, this prior structure leads to a closed-form complex-valued EMVS algorithm, referred to as C-EMVS here, that allows for fast identification of active voxels. Once again, note that the real-valued representation of this prior is given by

$$(\gamma_{j,Re}, \gamma_{j,Im})' \sim \left(1 - \psi_j^v\right) g_0^r\left(\left(\gamma_{j,Re}^v, \gamma_{j,Im}^v\right)'\right)$$
$$+ \psi_j^v g^r\left(\left(\gamma_{Re,j}^v, \gamma_{Im,j}^v\right)'\right)$$

with $g_0^r((\gamma_{Re,j}^v, \gamma_{Im,j}^v)') = N_2(\mathbf{0}, v_0\sigma_v^2 \mathbf{I}_2)$ and $g^r((\gamma_{Re,j}^v, \gamma_{Im,j}^v)') = N_2(\mathbf{0}, v_1\sigma_v^2 \mathbf{I}_2)$.

We complete the prior specification taking $\sigma_v^2 \sim IG(a_\sigma, b_\sigma)$, $\psi_j^v \sim \operatorname{Bernoulli}(\theta_j)$, with $\theta_j \sim \operatorname{Beta}(a_\theta, b_\theta)$, for all $j = 1, \ldots, p$ and $a_\sigma, b_\sigma, a_\theta, b_\theta$ constants. In particular, as discussed in the examples, we consider $a_\sigma = b_\sigma = 1/2$ and values of $v_0, v_1, a_\theta,$ and b_θ selected following guidelines similar to those provided in Rockova and George (2014) and Wang et al. (2015). This prior structure relates voxels through the common probability that the binary variables for a given task j are equal to one, that is, $\Pr(\psi_j^v = 1 \mid \theta_j) = \theta_j$, for all the voxels $v = 1, \ldots, V$.

2.2. Posterior Inference

We summarize the algorithms for posterior inference below. We first describe a complex-valued EMVS algorithm, C-EMVS, that leads to fast detection of active sites under the Bayesian model. A similar EMVS algorithm can be derived for magnitude-only models. We then provide a Markov chain Monte Carlo (MCMC) scheme that allows us to obtain full posterior inference. The simulations and experimental data analyzed in Sections 3 and 4 focus on the performance of the complex-valued and magnitude-only EMVS algorithms, as full MCMC is usually not

computationally efficient for the analysis of large-dimensional voxel-level fMRI and CV-fMRI.

2.2.1. A C-EMVS Algorithm for Fast Posterior Computations

Rockova and George (2014) proposed an expectation-maximization approach to Bayesian variable selection (EMVS) that takes advantage of the continuity of the spike distribution to produce rapidly computable closed-form expressions. Here, we develop an EMVS-based approach to posterior computation that combines the linear and complex-valued Gaussian structure in (4), the complex-valued spike-and-slab prior for $\boldsymbol{\gamma}^v$ in (7), and the priors for the remaining model parameters described in Section 2.1 above. More specifically, we now summarize the steps of the C-EMVS algorithm for the simplest model specification considered in the simulation studies presented in Section 3 (algorithms for general models are detailed in the online Appendices). This model is given by

$$\mathbf{y}^v = \mathbf{X}\boldsymbol{\gamma}^v + \boldsymbol{\eta}^v, \quad \boldsymbol{\eta}^v \sim CN_T\left(\mathbf{0}, 2\sigma_v^2 \mathbf{I}, \mathbf{0}\right),$$
$$\gamma_j^v \mid \psi_j^v \sim \left(1 - \psi_j^v\right) CN_1\left(0, 2v_0\sigma_v^2, 0\right)$$
$$+ \psi_j^v CN_1\left(0, 2v_1\sigma_v^2, 0\right), \quad j = 1, \ldots, p, \qquad (8)$$
$$\sigma_v^2 \sim IG\left(a_\sigma, b_\sigma\right), \quad \psi_j^v \mid \theta_j \sim \operatorname{Bernoulli}\left(\theta_j\right),$$
$$\theta_j \sim \operatorname{Beta}\left(a_\theta, b_\theta\right).$$

Note that, for each task j, model (8) relates voxels through the common probability that the binary variables that specify the activation at the voxel-level for such task are equal to one, that is, $\Pr(\psi_j^v = 1 \mid \theta_j) = \theta_j$ for all voxels $v = 1, \ldots, V$ and each task $j = 1, \ldots, p$. Letting $\boldsymbol{\gamma} = [\boldsymbol{\gamma}^1, \ldots, \boldsymbol{\gamma}^V]$, $\boldsymbol{\psi} = [\boldsymbol{\psi}^1, \ldots, \boldsymbol{\psi}^V]$, with $\boldsymbol{\psi}^v = (\psi_1^v, \ldots, \psi_p^v)'$, $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_p)'$, $\boldsymbol{\sigma}^2 = [\sigma_1^2, \ldots, \sigma_V^2]$, and $\mathbf{y} = [\mathbf{y}^1, \ldots, \mathbf{y}^V]$, we find that the full posterior density is given by

$$\begin{aligned} \pi\left(\boldsymbol{\gamma}, \boldsymbol{\psi}, \boldsymbol{\theta}, \boldsymbol{\sigma}^2 \mid \mathbf{y}\right) &\propto \prod_{v=1}^{V}\left[f\left(\mathbf{y}^v \operatorname{mid} \boldsymbol{\gamma}^v, \sigma_v^2\right) \pi\left(\boldsymbol{\gamma}^v \mid \boldsymbol{\psi}^v, \sigma_v^2\right)\right. \\ &\quad \left. \times \pi\left(\boldsymbol{\psi}^v \mid \boldsymbol{\theta}\right) \pi\left(\sigma_v^2\right)\right] \pi\left(\boldsymbol{\theta}\right) \\ &\propto \prod_{v=1}^{V}\left[CN_T\left(\mathbf{y}^v \mid \mathbf{X}\boldsymbol{\gamma}^v, 2\sigma_v^2 \mathbf{I}, \mathbf{0}\right)\right. \\ &\quad \left. \times CN_p\left(\boldsymbol{\gamma}^v \mid \mathbf{0}, \sigma_v^2 \boldsymbol{\Omega}_v, \mathbf{0}\right)\right] \\ &\quad \times \prod_{v=1}^{V}\left[\pi\left(\sigma_v^2\right) \prod_{j=1}^{p} \operatorname{Bernoulli}\left(\psi_j^v \mid \theta_j\right)\right] \\ &\quad \times \prod_{j=1}^{p} \operatorname{Beta}\left(\theta_j \mid a_\theta, b_\theta\right), \qquad (9) \end{aligned}$$

where $\boldsymbol{\Omega}_v = 2 \times \operatorname{diag}((1 - \psi_1^v)v_0 + \psi_1^v v_1, \ldots, (1 - \psi_p^v)v_0 + \psi_p^v v_1)$.

An EM algorithm for maximizing the full posterior $\pi(\boldsymbol{\gamma}, \boldsymbol{\theta}, \boldsymbol{\sigma}^2 \mid \mathbf{y})$ for this complex-valued model, referred to as C-EMVS, is derived by iteratively maximizing the objective function

$$\begin{aligned} Q(\boldsymbol{\gamma}, \boldsymbol{\theta}, \boldsymbol{\sigma}^2 \mid \boldsymbol{\gamma}^{(l)}, \boldsymbol{\theta}^{(l)}, \boldsymbol{\sigma}^{2,(l)}) \\ = E_{\boldsymbol{\psi}|\cdot}[\log \pi(\boldsymbol{\gamma}, \boldsymbol{\psi}, \boldsymbol{\theta}, \boldsymbol{\sigma}^2 \mid \mathbf{y}) \mid \boldsymbol{\gamma}^{(l)}, \boldsymbol{\theta}^{(l)}, \boldsymbol{\sigma}^{2,(l)}, \mathbf{y}], \end{aligned}$$

at iteration $l + 1$, where $E_{\boldsymbol{\psi}|\cdot}(\cdot) = E_{\boldsymbol{\psi}|\boldsymbol{\gamma}^{(l)},\boldsymbol{\theta}^{(l)},\sigma^{2,(l)},\mathbf{y}}(\cdot)$. Note that at iteration $l + 1$, the function $Q(\cdot)$ uses the maxima found at iteration l. Given the form of the log posterior in this case, we can write

$$Q(\boldsymbol{\gamma},\boldsymbol{\theta},\sigma^2 \mid \boldsymbol{\gamma}^{(l)},\boldsymbol{\theta}^{(l)},\sigma^{2,(l)})$$
$$= Q_1(\boldsymbol{\gamma},\sigma^2 \mid \boldsymbol{\gamma}^{(l)},\boldsymbol{\theta}^{(l)},\sigma^{2,(l)}) + Q_2(\boldsymbol{\theta} \mid \boldsymbol{\gamma}^{(l)},\boldsymbol{\theta}^{(l)},\sigma^{2,(l)}) + K_Q, \quad (10)$$

with $\quad Q_1(\boldsymbol{\gamma},\sigma^2 \mid \boldsymbol{\gamma}^{(l)},\boldsymbol{\theta}^{(l)},\sigma^{2,(l)}) = \sum_{v=1}^{V} Q_1^v(\boldsymbol{\gamma}^v,\sigma_v^2 \mid \boldsymbol{\gamma}^{v,(l)},\boldsymbol{\theta}^{(l)},\sigma_v^{2,(l)})$ and K_Q a constant. For the E-step, we compute the conditional expectations $E_{\psi^v|\cdot}[\psi_j^v]$ and $E_{\psi^v|\cdot}[\frac{1}{(1-\psi_j^v)v_0+\psi_j^v v_1}]$. The M-step solves for $(\boldsymbol{\gamma}^{(l+1)},\sigma^{2,(l+1)})$ and $\boldsymbol{\theta}^{(l+1)}$ by maximizing Q_1^v for $v = 1,\ldots,V$ and Q_2 in (10). The complete details for this C-EMVS algorithm, as well as those for algorithms under more general complex-valued priors (e.g., noncircular priors) can be found in the online Appendices.

The C-EMVS algorithm is iterated until $\|\boldsymbol{\gamma}^{(l)} - \boldsymbol{\gamma}^{(l-1)}\| < \epsilon$, $\|\boldsymbol{\theta}^{(l)} - \boldsymbol{\theta}^{(l-1)}\| < \epsilon$ and $\|\sigma^{2,(l)} - \sigma^{2,(l-1)}\| < \epsilon$, with ϵ small. In the analyses of simulated and human experimental data presented in Sections 3 and 4, we use $\epsilon = 10^{-3}$. We assess convergence by monitoring that the log-posterior distribution increases at each step of the algorithm. Once the EM algorithm converges, we obtain estimated posterior modes $\hat{\boldsymbol{\gamma}}$, $\hat{\sigma}^2$, and $\hat{\boldsymbol{\theta}}$. Then, for each voxel we compute $Pr(\psi_j^v = 1 \mid \hat{\boldsymbol{\gamma}},\hat{\boldsymbol{\theta}},\hat{\sigma}^2,\mathbf{y})$, and we label a given voxel v active for task j if $Pr(\psi_j^v = 1 \mid \hat{\boldsymbol{\gamma}},\hat{\boldsymbol{\theta}},\hat{\sigma}^2,\mathbf{y}) > \delta$, where δ is a fixed threshold value. This is equivalent to saying that a voxel is active if its corresponding strength is greater than some real-valued threshold $\gamma_j^{*,v}$, that is, $|\hat{\gamma}_j^v| > \gamma_j^{*,v}$. A common choice of δ is 0.5, which leads to a local version of the median probability model of Barbieri and Berger (2004). Some researchers in the fMRI community suggest using $\delta = 0.8722$ for magnitude-only models. Smith and Fahrmeir (2007) gave a clear description of the motivation for this threshold value in the context of a Bayesian spatial model. Given that our models do not explicitly incorporate a spatial structure, we use $\delta = 0.5$ in the following analyses. A further alternative that could be considered within a Bayesian decision-theoretical framework is to choose the threshold by minimizing a well-defined loss function, or via Bayesian false discovery rates (see, e.g., Müeller, Parmigiani, and Rice, 2006 and Sun et al., 2015).

Finally note that, if desired, the algorithm can also be implemented for the real-valued version of the model in (8) given by

$$\begin{pmatrix} \mathbf{y}_{Re}^v \\ \mathbf{y}_{Im}^v \end{pmatrix} = \mathbf{X}^r \boldsymbol{\gamma}_r^v + \boldsymbol{\eta}_r^v, \quad \boldsymbol{\eta}_r^v \sim N_{2T}\left(\mathbf{0},\sigma_v^2\mathbf{I}_{2T}\right),$$

$$\boldsymbol{\gamma}_r^v = \begin{pmatrix} \boldsymbol{\gamma}_{Re}^v \\ \boldsymbol{\gamma}_{Im}^v \end{pmatrix} \sim N_{2p}\left(\mathbf{0},\sigma_v^2\begin{pmatrix} \boldsymbol{\Sigma}_{Re,Re}(\boldsymbol{\psi}^v) & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_{Im,Im}(\boldsymbol{\psi}^v) \end{pmatrix}\right)\right),$$

and the same priors on σ_v^2, ψ_j^v, and θ_j specified above.

2.2.2. Posterior Inference via Markov Chain Monte Carlo

Full posterior inference can be obtained via MCMC. Similar to the C-EMVS case described above, we generalize the Stochastic Search Variable Selection algorithm (SSVS) proposed by George and McCulloch (1993) to the complex-valued domain. Suppose we have a simplified complex-valued model such as (8) except

that we now use a "nonconjugate" version of the spike-and-slab prior on $\boldsymbol{\gamma}^v$, that is, $\gamma_j^v \mid \psi_j^v \sim (1-\psi_j^v)CN_1(0,2v_0,0) + \psi_j^v CN_1(0,2v_1,0)$, $j = 1,\ldots,p$. The general vectorized form of this prior can be written as $\boldsymbol{\gamma}^v \mid \boldsymbol{\psi}^v \sim CN_p(\mathbf{0},\boldsymbol{\Omega}_v,\mathbf{0})$, with $\boldsymbol{\Omega}_v = 2 \times \text{diag}[(1-\psi_1^v)v_0 + \psi_1^v v_1,\ldots,(1-\psi_p^v)v_0 + \psi_p^v v_1]$. Then, the posterior full conditional distributions for a Gibbs sampling scheme can be derived as follows:

- For each v, $v = 1,\ldots,V$, $\boldsymbol{\gamma}^v \mid \mathbf{y}^v,\sigma_v^2,\boldsymbol{\psi}^v \sim CN_p(\boldsymbol{\mu}_v^v,\boldsymbol{\Omega}_{\text{pos}}^v,\mathbf{0})$, with $\boldsymbol{\Omega}_{\text{pos}}^v = (2^{-1}\sigma_v^{-2}\mathbf{X}'\mathbf{X} + \boldsymbol{\Omega}_v^{-1})^{-1}$, and $\boldsymbol{\mu}_{\boldsymbol{\gamma}}^v = \boldsymbol{\Omega}_{\text{pos}}^v\mathbf{X}'\mathbf{y}^v/\sigma_v^2$.
- $\sigma_v^2 \mid \mathbf{y}^v,\boldsymbol{\gamma}^v \sim IG(a_\sigma^{v,\text{pos}},b_\sigma^{v,\text{pos}})$, with $a_\sigma^{v,\text{pos}} = T + a_\sigma$ and $b_\sigma^{v,\text{pos}} = \|\mathbf{y}^v - \mathbf{X}\boldsymbol{\gamma}^v\|^2/2 + b_\sigma$.
- $Pr(\psi_j^v = 1 \mid \mathbf{y}^v,\boldsymbol{\gamma}^v,\sigma^2,\theta,\boldsymbol{\psi}_{-j}^v) = \frac{c_j^v}{c_j^v + e_j^v}$, with $c_j^v = \pi(\boldsymbol{\gamma}_j^v \mid \mathbf{y}^v,\psi_j^v = 1,\boldsymbol{\psi}_{-j}^v) \times \theta_j$ and $e_j^v = \pi(\boldsymbol{\gamma}_j^v \mid \mathbf{y}^v,\psi_j^v = 0,\boldsymbol{\psi}_{-j}^v) \times (1-\theta_j)$. Here $\pi(\boldsymbol{\gamma}_j^v \mid \mathbf{y}^v,\psi_j^v = 1,\boldsymbol{\psi}_{-j}^v)$ and $\pi(\boldsymbol{\gamma}_j^v \mid \mathbf{y}^v,\psi_j^v = 0,\boldsymbol{\psi}_{-j}^v)$ are complex-normal densities (see online Appendices for details).
- For each j, $j = 1,\ldots,p$, $\theta_j \mid \mathbf{y},\boldsymbol{\psi}^v \sim \text{Beta}(\sum_{v=1}^{V}\psi_j^v + a_\theta, V - \sum_{v=1}^{V}\psi_j^v + b_\theta)$.

To decide whether a voxel v is active or not after MCMC convergence is achieved, we look at the posterior probability of $\psi_j^v = 1$, for each task-related BOLD signal $j = 1,\ldots,p$. A detailed derivation of general complex-valued SSVS algorithm and the corresponding full conditional distributions above can be found in the online Appendices.

3. Simulation Studies

We show the performance of the proposed complex-valued variable selection methods for detecting activation in two simulation studies. The first study compares the C-EMVS algorithm to computationally fast alternatives that are often used in practice, such as lasso and adaptive lasso (Tibshirani 1996; Zou 2006). We also compare the results obtained by the proposed complex-valued model and priors via the C-EMVS algorithm with those obtained using a magnitude-only Bayesian model with the real-valued priors in Rockova and George (2014). The magnitude-only voxel time series courses are obtained by taking the moduli of the CV-fMRI signals at each voxel. The second study considers a physically realistic simulated CV-fMRI dataset.

3.1. Simulation Study I

We simulated 20 datasets consisting of 48×48 CV-fMRI slices with a constant baseline signal and a single expected BOLD signal (i.e., $p = 1$) resulting from the convolution of a stimulus indicator function and the canonical hemodynamic response function. Three activation regions were simulated using the function `specifyregion` in the R package `neuRosim` (Welvaert et al. 2011). More specifically, for $v = 1,\ldots,48 \times 48$, and $t = 1,\ldots,200$, the time series for each voxel v were simulated as follows:

$$y_{t,Re}^v = (\beta_0 + \beta_1 f_v z_t)\cos(\alpha_0) + \eta_{t,Re}^v, \quad \eta_{t,Re}^v \sim N(0,\sigma^2)$$
$$y_{t,Im}^v = (\beta_0 + \beta_1 f_v z_t)\sin(\alpha_0) + \eta_{t,Im}^v, \quad \eta_{t,Im}^v \sim N(0,\sigma^2), \quad (11)$$

where f_v is the BOLD signal strength or intensity rate of voxel v, with $f_v = 0$ if voxel v is nonactive and $f_v \neq 0$ if voxel v is

Experimental Design

BOLD signal with canonical HRF

f_v
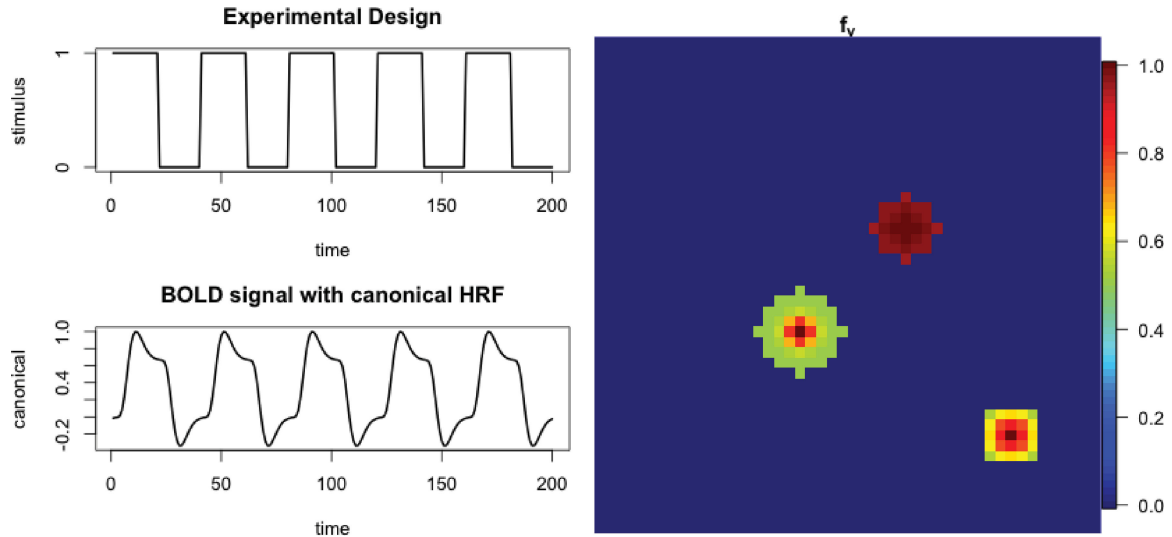
Figure 1. Left: Block experimental design (top); expected BOLD signal obtained from convolving the stimulus indicator signal with the canonical hemodynamic function (bottom). Right: Activation regions and f_v values for active voxels.

in an active region. The values of f_v for active voxels were specified using the argument `fading` in the function `specifyregion` in `neurosim`. Here, the fading of the expected BOLD signal decays exponentially depending on the distance of the active voxel v with coordinates (i, j), to the center of the active region with coordinates (i', j'), that is, the fading for voxel v is given by

$$f_v(i, j) = \frac{1}{4}\{2 \cdot \exp[-((i - i')^2 + (j - j')^2) \cdot \varrho] + 2\},$$

where ϱ is the decay rate in $[0, 1]$ with 0 and 1 corresponding, respectively, to no decay and to the strongest decay. z_t in (11) is the BOLD signal given by the convolution of the canonical HRF, denoted as h_t, and the stimulus indicator function s_t, that is, $z_t = h_t \otimes s_t$.

We used $\alpha_0 = \pi/4$ and different values of β_0 and β_1 to examine the performance of the proposed complex-valued models using the C-EMVS algorithm for posterior computations. These were chosen to set specific values of the SNR and the contrast-to-noise ratio (CNR) as defined in Rowe and Logan (2004), with $\text{SNR} = \beta_0/\sigma$ and $\text{CNR} = \beta_1/\sigma$. Note that active voxels have different CNRs given by $\text{CNR}_v = (\beta_1 f_v)/\sigma$, with $\text{CNR}_v \leq \text{CNR}$ for all v, as $f_v \in [0, 1]$. Hence, the largest CNR for active voxels is β_1/σ, computed using no fading, while the smallest CNR is $\beta_1 f_{\min}/\sigma$, where $f_{\min} = \min_{\{v \in A\}} f_v$ and A is the set of active voxels. The average CNR is $\sum_{\{v \in A\}} \beta_1 f_v/|A|$. In this simulation, we used $f_{\min} \approx 0.50$ and $\sum_{\{v \in A\}} f_v/|A| \approx 0.71$, with $|A| = 103$ active voxels, which accounts for 4.47% of all voxels.

The top left plot in Figure 1 shows the experimental block design, with $s_t = 1$ if the stimulus is on and $s_t = 0$ otherwise. It consists of five epochs of 20 sec on and 20 sec off with an observation interval of 1. The resulting BOLD signal z_t is shown in the bottom left plot. The right plot displays the active regions with the corresponding f_v values. The three active regions are centered at the coordinates $(20, 20)$, $(30, 30)$, and $(40, 10)$, with radius arguments 3, 2, 1, and fading arguments 0.5, 0.01, and 0.3, respectively, for each region. The bottom-right region is a square and the other two are circles.
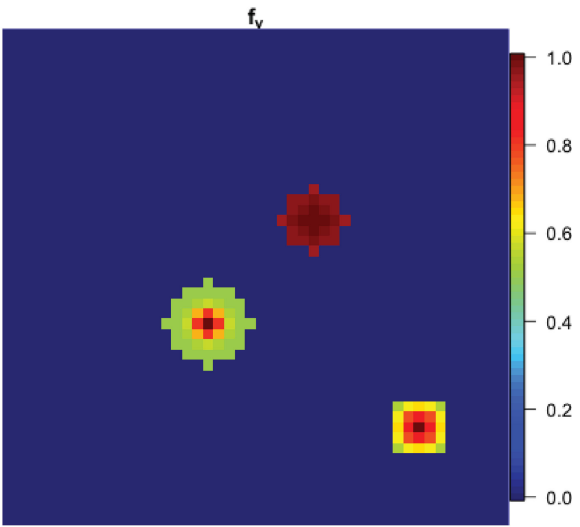
Four different SNRs, 0.5, 1, 5, and 10, and three different CNRs, 0.5, 1, and 1.5, were considered, resulting in 12 different SNR-CNR data types. These are numbered as shown in Table 1. We generated 20 simulated datasets for each SNR-CNR data type and computed classification performance measures (sensitivity, specificity, precision, and accuracy) to examine how well our algorithm and other methods perform in the different scenarios.

Four methods are compared in this simulation study, the proposed Bayesian complex-valued model using the C-EMVS algorithm for posterior computations (referred to as CV in the results below), the Bayesian magnitude-only model with the EMVS algorithm (MO), and the lasso (LA) and adaptive lasso (ALA), both for magnitude-only data.

The Bayesian complex-valued model used here has the form

$$y_t^v = \gamma_1^* + \gamma_2^{*,v} x_t + \eta_t^v, \quad \eta_t^v \sim \text{CN}_1\left(0, 2\sigma^2, 0\right),$$

with γ_1^* a baseline parameter and $\gamma_2^{*,v}$ the complex-valued activation parameters for each voxel and $x_t = z_t$. For the baseline parameter, we use a prior of the form $\pi(\gamma_1^*) \propto 1$. For the activation parameters and the remaining model parameters, we used the following priors:

$$\gamma_2^{*,v} \mid \psi^v \sim (1 - \psi^v) \text{CN}_1\left(0, 2v_0\sigma^2, 0\right) + \psi^v \text{CN}_1\left(0, 2v_1\sigma^2, 0\right),$$
$$\sigma^2 \sim \text{IG}(1/2, 1/2), \quad \psi^v \mid \theta \sim \text{Bernoulli}(\theta),$$
$$\theta \sim \text{Beta}(1, 1). \tag{12}$$

The baseline parameter was integrated out before proceeding with the C-EMVS or MCMC algorithms for posterior inference and detection of active sites, so we used the algorithms outlined in Section 2 and detailed in the online Appendices.

We also consider a Bayesian model for the magnitude-only data. The magnitude-only time courses are obtained as $y_{t,\text{Mag}}^v = \sqrt{(y_{t,Re}^v)^2 + (y_{t,Im}^v)^2}$. The MO model used to analyze these data is

Table 1. Twelve data types and their corresponding SNR and CNR.

SNR	0.5			1			5			10		
CNR	0.5	1	1.5	0.5	1	1.5	0.5	1	1.5	0.5	1	1.5
Data type	1	2	3	4	5	6	7	8	9	10	11	12
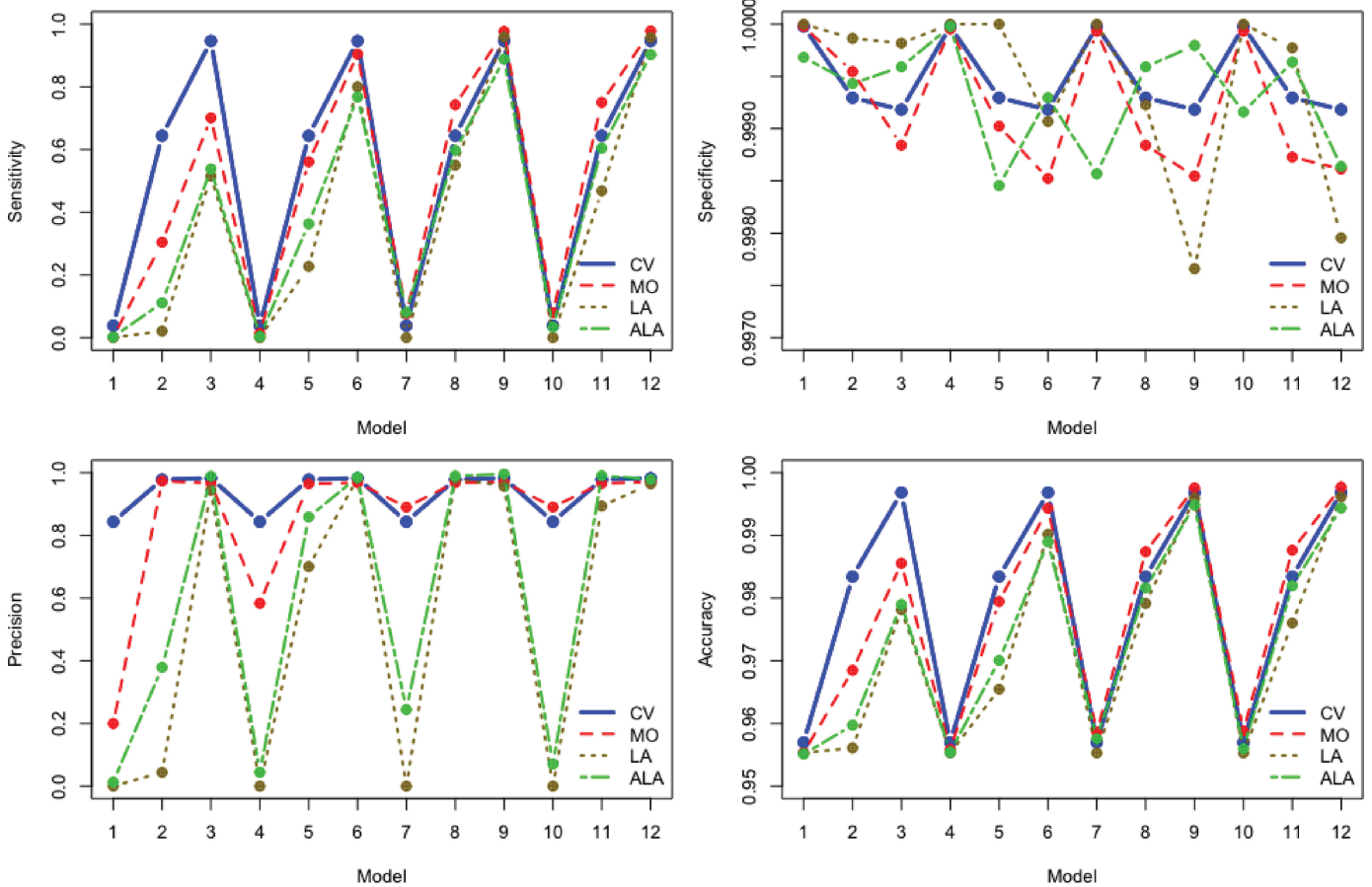
Figure 2. Sensitivity (top-left), specificity (top-right), precision (bottom-left), and accuracy (bottom-right) for four models: Complex-valued EM (CV; blue, solid), magnitude-only EM (MO; red, dash), Lasso (LA; brown, dotted), and adaptive Lasso (ALA; green, dash-dotted).

essentially the same as the CV model used for the complex-valued data, except that the linear model is now real-valued and the priors on the regression coefficients are real-valued Gaussian spike-and-slab priors. This is

$$y_{t,\mathrm{Mag}}^{v} = \gamma_{\mathrm{Mag},1}^{*} + \gamma_{\mathrm{Mag},2}^{*,v} x_t + \eta_t^{v}, \quad \eta_t^{v} \sim N\left(0, \sigma^2\right),$$

$$\gamma_{\mathrm{Mag},2}^{*,v}|\psi_{\mathrm{Mag}}^{v} \sim \left(1 - \psi_{\mathrm{Mag}}^{v}\right) N_1\left(0, v_0\sigma^2\right) + \psi_{\mathrm{Mag}}^{v} N_1\left(0, v_1\sigma^2\right),$$

$$\sigma^2 \sim \mathrm{IG}\left(1/2, 1/2\right), \quad \psi_{\mathrm{Mag}}^{v}|\theta \sim \mathrm{Bernoulli}\left(\theta\right),$$

$$\theta \sim \mathrm{Beta}(1, 1),$$

and $\pi\left(\gamma_{\mathrm{Mag},1}^{*}\right) \propto 1$.

The tuning parameters in the Bayesian CV and MO models above, v_0 and v_1, are chosen as suggested in Rockova and George (2014) and Wang et al. (2015). More specifically, we fix v_1, taking $v_1 = 1$ and choose the optimal v_0 in each case, denoted as v_0^{CV} and v_0^{MO}, for the CV and MO models, respectively, by maximizing the marginal posterior $\pi_0(\boldsymbol{\psi} \mid \mathbf{y})$ that evaluates $\boldsymbol{\psi}$ according to the submodel that contains only those variables for which $\psi_j^{v} = 1$. This marginal can be derived in closed form up to a normalizing constant. From our experience with the real and simulated datasets analyzed here, the optimal v_0 takes values around $1/\sqrt{100Tp}$ and usually lies in the interval $(1/\sqrt{1000Tp}, 1/\sqrt{10Tp})$, where p is the number of tasks. In this simulation, we only have one task so $p = 1$.

Finally, we also applied the lasso (LA) and adaptive lasso (ALA) methods (Tibshirani 1996; Zou 2006) to the magnitude-only data. Both LA and ALA use a regularization parameter and ALA uses additional weights to allow for different penalizations in the regression coefficients (the $\gamma_{\mathrm{Mag},2}^{*,v}$ parameters in our case). The regularization parameter was chosen using a five-fold cross-validation approach and the weights in the ALA were set to $1/|\hat{\gamma}_{\mathrm{Mag},2}^{*,v}|$, where $\hat{\gamma}_{\mathrm{Mag},2}^{*,v}$ is the ordinary least-square estimator of $\gamma_{\mathrm{Mag},2}^{*,v}$. LA and ALA were implemented using the R package `glmnet` (Friedman, Hastie, and Tibshirani 2010).

The resulting average performance measures over the 20 simulated datasets for the four different methods are summarized in Figure 2. Note that this simulation contains 2201 nonactive voxels out of a total 2304 voxels, so any model can achieve 95.53% accuracy by simply classifying all voxels as nonactive. Hence, the accuracy subfigure is plotted from 0.95 to 1 for clear comparison. Similarly, the specificity is plotted from 0.997 to 1.

First, we see that both Bayesian variable selection approaches, the one for the CV-fMRI and the one for magnitude-only data (MO), dominate the traditional lasso (LA) and adaptive lasso (ALA) for magnitude-only data in terms of sensitivity (power), precision, and accuracy. The Bayesian approaches are able to eliminate most of the false positives by borrowing strength across voxels via the common probability of activation parameter θ. The Bayesian CV and MO methods are comparable to lasso and adaptive lasso in terms of specificity, while the first provide a more complete inferential analysis. The main

(a) Activation maps
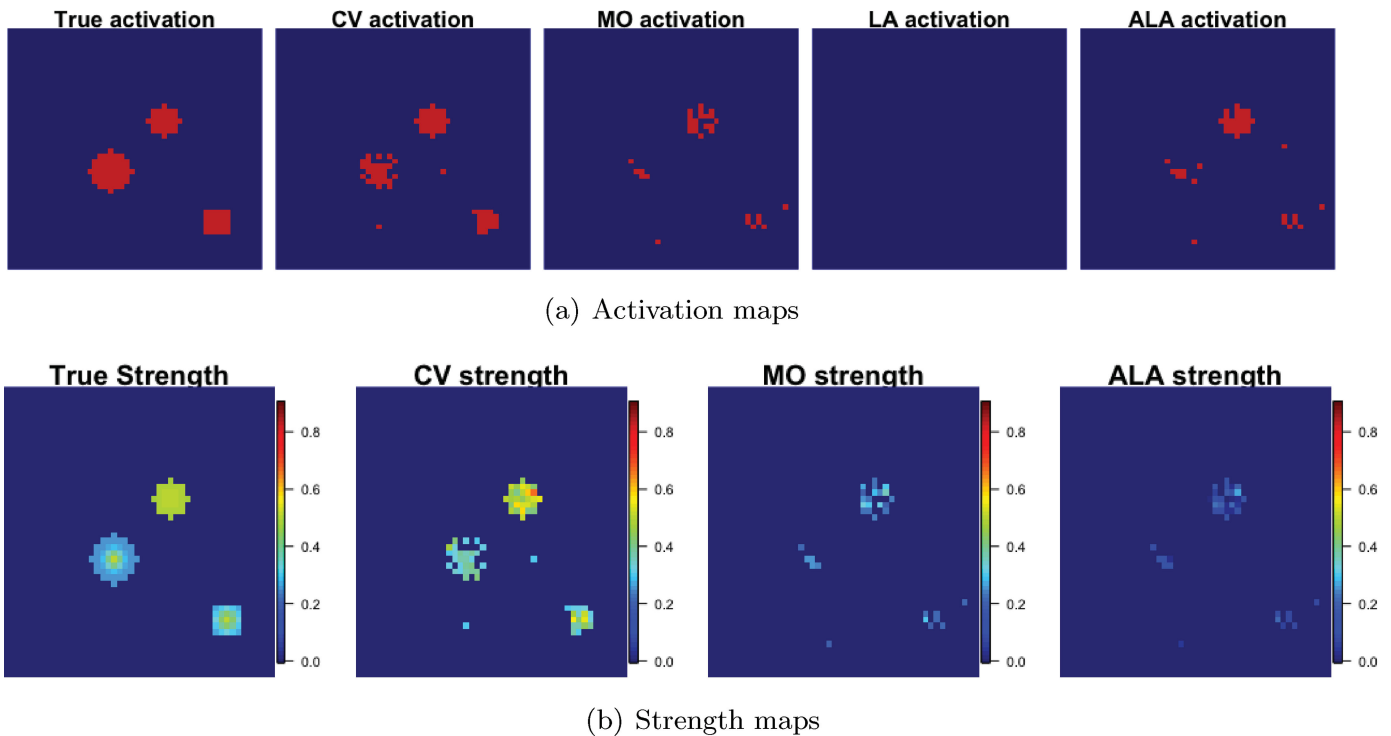
(b) Strength maps

Figure 3. Activation and strength maps for a simulated dataset with SNR $= 0.5$ and CNR $= 1$. (a) Activation maps showing the true active sites, and the activation results obtained from C-EMVS, MO-EMVS, Lasso, and Adaptive Lasso. Activated sites are colored in red. (b) Strength maps: true strength and estimated strengths from C-EMVS, MO-EMVS, Lasso, and adaptive Lasso.

advantage of the Bayesian CV model with respect to the Bayesian MO model is that the CV model significantly detects more true positives than the MO when the SNR is small, which leads to higher sensitivity, precision, and accuracy. When the SNR is fairly large, using the information provided only by the magnitude leads to good activation results in these simulated scenarios. In fact, the MO model even has a slightly larger sensitivity than the CV model when the SNR is 5 or 10. On the other hand, the CV model leads to higher specificity and precision than the MO model even when the SNR is 5 or 10. Moreover, the performance of the CV model is very consistent across different SNRs. Hence, when the CV-fMRI data are recorded under small SNRs or when researchers are uncertain about the magnitude of the SNR in their data, the CV model stands out as the best option among the models considered here. Given that improved MRI technology allows for improved spatial resolution and therefore reduces SNR, we would expect that complex-valued models will become an essential tool for detecting active sites in CV-fMRI data.

Figure 3 shows the true activation and strength maps for one of the 20 simulated datasets with SNR = 0.5 and CNR = 1 along with the estimated activation and strength maps (only for sites labeled as active) obtained from the C-EMVS (CV), the magnitude-only EMVS (MO), and adaptive lasso (ALA). The strength maps for lasso are not shown, as lasso detected no active sites. Both activation maps for the complex-valued and magnitude-only EMVS display activation levels that result from setting $v_1 = 1$ and choosing the optimal values of v_0 for each method as discussed above. For this dataset and with our prior distribution settings, we found that the optimal values were $v_0^{CV} = 0.0071$ and $v_0^{MO} = 0.0056$. The C-EMVS approach clearly outperforms all the other approaches: it has higher power

for detecting active voxels while simultaneously controlling for false positives, and also leads to more accurate estimation of the activation strength (note that MO and ALA clearly underestimate the strength). In relation to this point, we computed the mean squared errors (MSEs) for this simulated dataset under the C-EMVS, MO, and ALA approaches for voxels that are labeled as active for at least one of the three methods and found that the MSEs values were, respectively, 0.0080, 0.0084, and 0.1162. The complex-valued model also leads to more accurate inference for σ. Magnitude-only models underestimate σ when the SNR is small as a consequence of the fact that the MO error distribution is truly Ricean at low SNRs. This can lead to an increase of false positives when detecting activation (in fact, we can see that the specificity values obtained with the complex-valued model are generally higher than those obtained with magnitude-only model as shown in Figure 2). For example, for a dataset generated under a true value of $\sigma = 0.5$, when SNR = 0.5, we found $\hat{\sigma}_{CV} = 0.497$, while $\hat{\sigma}_{MO} = 0.346$. To obtain better estimates of σ with MO models, we need to considerably increase the SNR. For instance, for a simulated dataset with SNR = 10, we obtained $\hat{\sigma}_{MO} = 0.495$ which is closer to the true value 0.5. These results are consistent with the findings of Rowe (2005b).

Finally, we also implemented the MCMC sampling approach outlined in Section 2 and detailed in the online Appendices to achieve full posterior inference for the complex-valued models. We obtained similar results to those from the C-EMVS algorithm in terms of the number of active sites and the strength of those sites, but we highlight that, in addition, the MCMC approach allows us to compute uncertainty measures related to activation strength and any other functions of the model parameters. For instance, Figure 4 shows posterior mean strength maps and 95% posterior credibility strength maps for a single dataset
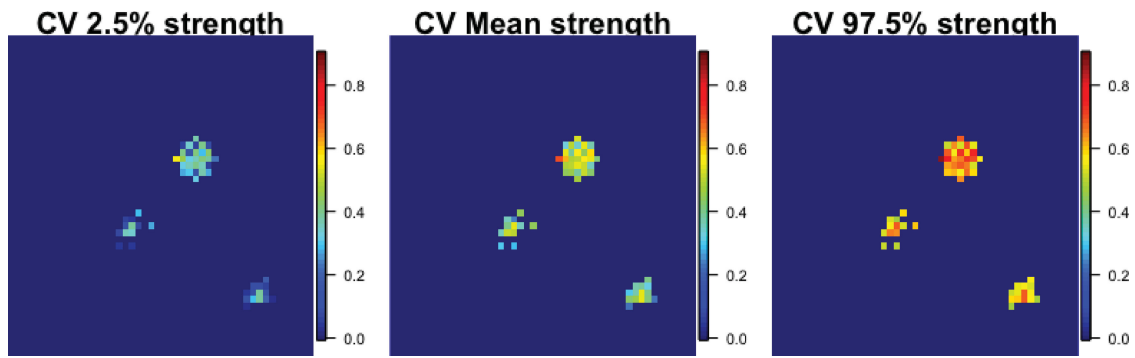
Figure 4. Strength maps for a simulated dataset with SNR = 0.5 and CNR = 1 obtained from a complex-valued model via MCMC. Left: 2.5% quantile map; Middle: Posterior mean map; Right: 97.5% quantile map.

obtained from the complex-valued model. As seen in this figure, the posterior mean estimates for the strength are similar to those obtained via the C-EMVS algorithm but the MCMC-based posterior credibility maps provide additional information about the strength maps. We see that, in general, there is less uncertainty about activation strength for voxels located in region centered at (30,30) than for voxels located in the region centered at (40,10). This makes sense given the true strength maps used to generate the simulated data (see Figure 3). In cases where this Gibbs sampling scheme is not computationally feasible (e.g., when several large-dimensional images for multiple subjects need to be analyzed), one could consider a hybrid approach that, say, uses the C-EMVS method to determine which sites are active and then uses the Gibbs sampling scheme only on regions of the brain that present active sites to obtain posterior uncertainty measures on strength maps and/or activation maps for those regions only. Alternative methods based on obtaining approximate inference via variational Bayes could also be considered (see, e.g., Yu et al., 2016).

3.1.1. Additional Structure: Temporal Correlation

We also analyzed synthetic CV-fMRI data simulated under (11) but with errors following an autoregressive structure of order one, that is, $\eta_{t,Re}^v = \varphi \eta_{t-1,Re}^v + \zeta_{t,Re}^v$, with $\zeta_{t,Re}^v$ independent Gaussian for all t, $\zeta_{t,Re}^v \sim N\left(0, \sigma^2\right)$, and $\eta_{t,Im}^v = \varphi \eta_{t-1,Im}^v + \zeta_{t,Im}^v$, with $\zeta_{t,Im}^v$ also independent Gaussian for all t, $\zeta_{t,Im}^v \sim N(0, \sigma^2)$ and $\varphi \in [0, 1)$ the AR coefficient. We considered values of φ ranging from 0.1 to 0.9, and the same 12 SNR-CNR scenarios described in the previous simulation, with $\sigma^2 = 0.25$. We analyzed these data using two versions of the model $y_t^v = \gamma_1^* + \gamma_2^{*,v} x_t + \eta_t^v$: one version with η_t^v iid complex normal, and another version with η_t^v following a complex-valued AR(1) structure in $\eta_t^v = \eta_{t,Re}^v + i\eta_{t,Im}^v$ as described above. Figure 5 displays the sensitivity, specificity, precision, and accuracy for the two versions of the CV model (independent and autoregressive errors) and two types of data (AR errors with $\varphi = 0.5$ and $\varphi = 0.9$). Overall we find that the larger the value of φ the harder it is to detect active sites, particularly for small SNR and CNR. This makes sense, as AR(1) errors with φ close
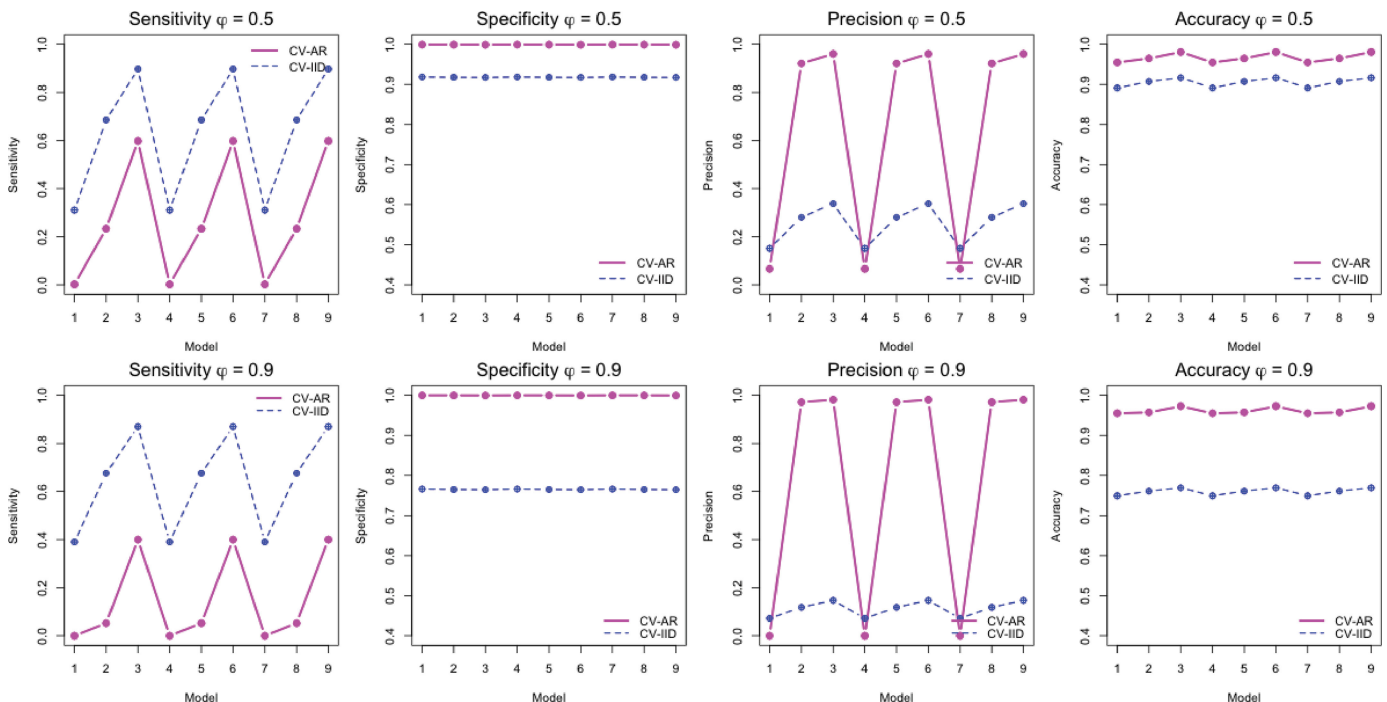
Figure 5. Sensitivity, specificity, precision, and accuracy plots for synthetic AR(1) CV-fMRI data with AR coefficients 0.5 (top plots) and 0.9 (bottom plots). The plots are based on results obtained from analyzing 20 datasets using models that assumed independent errors (dotted lines) and AR(1) errors (solid lines).

to 1 may add a temporal structure that locally resembles a linear trend and can easily hide/mask the temporal behavior that characterizes active sites due to increased variability in the observed time series. We also see that while the CV model with independent errors has higher sensitivity, it also leads to a larger number of false positives (we only have about 77% specificity for the model with independent errors while we obtain 100% specificity for the model with AR errors when $\varphi = 0.9$). Therefore, the CV model with AR errors is overall a better option in terms of specificity, precision, and accuracy, particularly when φ is large.

3.1.2. Additional Structure: HRF Effect and Prior Sensitivity Analyses

We also studied the effects of the HRF choice and the prior distributions. Regarding the HRFs, we analyzed the simulated and human data presented in Sections 3 and 4 with three different classes of HRF functions, namely, canonical, gamma, and boxcar with different choices for the parameters that define each particular class. For a given HRF, we can select the optimal v_0 and then choose the HRF and corresponding v_0 that leads to the smallest MSE (mean squared error) for a particular dataset. Overall we found that the MSEs for the optimal HRFs within each class were comparable. Furthermore, the results in terms of the number and locations of the sites labeled as active were also similar across the optimal HRFs within each class.

We studied the sensitivity of our posterior results with respect to the prior distributions. In particular, as mentioned above, we generally assume $\theta \sim \text{Beta}(1, 1)$. In cases where a sparser structure is desired a priori, that is, when it makes sense biologically to assume that the number of active sites is just a very small percentage of the total number of sites, priors of the form $\theta \sim \text{Beta}(1, b)$ with b large can be used. In this simulation study, we found that the activation results were essentially the same for any prior with $b \leq 1000$. Priors with values of $b > 1000$ lead to sparser results (i.e., less active sites) in the simulated data. For the human data presented in Section 4, we found that we are able to detect similar numbers and locations of active sites for priors with values of $b \in [1, 100,000]$. Note that choosing $b = 1000$ leads to a fairly informative prior, with about 0.09% of active sites expected a priori and rarely above 0.4% of active sites expected a priori.

Finally, we assessed the effect of using noncircular complex-normal priors on γ^v, that is, priors of the form $\gamma^v \mid \psi^v \sim \text{CN}_p(\mathbf{0}, \sigma_v^2 \mathbf{\Omega}_v, \sigma_v^2 \mathbf{\Lambda}_v)$, with $\mathbf{\Lambda}_v \neq \mathbf{0}$, so that there is a nonzero correlation between the real and imaginary components of γ^v. As expected, allowing for a correlation structure between the real and imaginary components of γ^v leads to improved results when such underlying structure is present in the data, that is, having a more flexible prior that accounts for this correlation leads to higher power for detecting activation and reduces the number of false positives. On the other hand, such priors also lead to models that are more computationally costly and may potentially lead to biases in the posterior results. Therefore, we recommend the use of noncircular priors only when there is a strong indication that there is a significant correlation between the real and complex components of γ^v, and that such correlation structure is similar for active and nonactive voxels. Alternative priors will be developed and investigated in the future but are out of the scope of this work.
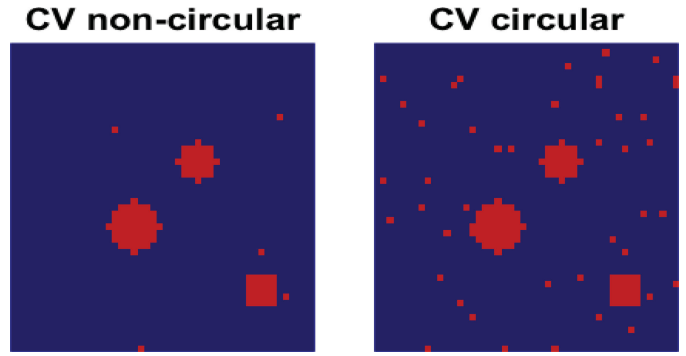
Figure 6. Left: Activation results obtained from a model with a noncircular prior. Right: Activation results obtained with a circular prior. The data were simulated so that the real and complex components of the activation coefficients are highly correlated.

We now illustrate the use of noncircular priors in the analysis of a simulated dataset with high correlation among the real and imaginary components for both types of voxels, active and nonactive. The data were simulated following:

$$y_{t,Re}^v = \left(\beta_0 + \beta_{1,Re}^v z_t\right) \cos(\alpha_0) + \eta_{t,Re}^v, \quad \eta_{t,Re}^v \sim N(0, \sigma^2),$$
$$y_{t,Im}^v = \left(\beta_0 + \beta_{1,Im}^v z_t\right) \sin(\alpha_0) + \eta_{t,Im}^v, \quad \eta_{t,Im}^v \sim N(0, \sigma^2),$$

with $\alpha_0 = \pi/4$, $\sigma^2 = 0.1$, SNR $= 0.4$, $\beta_0 = 0.8$, and the same z_t used in the previous simulation study. In addition, the parameters $\beta_{1,Re}^v$ and $\beta_{1,Im}^v$ were obtained from complex-normal distributions as follows. For active voxels, we sampled $\beta_{1,Re}^v + i\beta_{1,Im}^v$ from a complex noncircular normal with mean 0.7 and covariance and relation values that lead to a correlation of 0.9 between $\beta_{1,Re}^v$ and $\beta_{1,Im}^v$. For nonactive voxels, we sampled $\beta_{1,Re}^v + i\beta_{1,Im}^v$ from a complex noncircular normal with mean 0 and covariance and relation values that lead to a correlation of 0.9 between $\beta_{1,Re}^v$ and $\beta_{1,Im}^v$. Note that $\eta_{t,Re}^v$ and $\eta_{t,Im}^v$ are assumed independent for all the voxels and also across time. The location of the active voxels was determined using the same activation map used in the previous simulation and displayed in the left plot of Figure 3(a).

Figure 6 shows the results obtained from a model that uses a noncircular prior on γ^v that captures the induced correlation structure in these coefficients (left plot) and also shows the results obtained using a circular prior that assumes no correlation structure. Clearly, the model with a noncircular prior leads to much better results as it adequately identifies the active regions and leads to a much smaller number of false positives than those obtained under the model with the circular prior. The model with the noncircular prior also leads to better results in terms of estimation of activation strength and reduced MSE.

3.2. Simulation Study II: Physically Realistic Simulated Data

A more realistic simulated dataset was generated using a discrete version of the magnetic resonance (MR) signal equation after steady-state magnetization (Karaman, Bruce, and Rowe 2015). This equation is given by

$$s(k_x, k_y \mid t) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \rho(x, y) e^{-t/T_2^*(x,y)}$$
$$\times \left(1 - e^{-TR/T_1(x,y)}\right) e^{i\Gamma_H \Delta B(x,y)t} e^{-i2\pi(k_x x + k_y y)} dx dy,$$
$$\tag{13}$$

Table 2. Tissue physical parameter values.

	Gray matter	White matter	Cerebrospinal fluid	Outside brain
PSD	0.83	0.71	1.00	10^{-9}
T_2^*	59.7	54.6	2200	10^{10}
T_1	1331	832	4000	10^{-6}

where $s(k_x, k_y|t)$ is the k-space location at intra slice time t, $\rho(x, y)$ is the proton spin density (PSD), $T_2^*(x, y)$ is the transverse relaxation rate (TRR), $T_1(x, y)$ is the longitudinal relaxation rate (LRR), $\Delta B(x, y)$ is the magnetic field inhomogeneity (MFI), and Γ_H is the proton gyromagnetic ratio (Haacke et al. 1999). The k-space points in (13) are defined by the temporal integral of the magnetic field gradients $G_x(\cdot)$ and $G_y(\cdot)$:

$$k_x = \frac{\Gamma_H}{2\pi} \int_0^t G_x(t')dt', \quad \text{and} \quad k_y = \frac{\Gamma_H}{2\pi} \int_0^t G_y(t')dt'.$$

As input to this data-generation process, 3.0 T tissue specific physical parameters (Peters et al. 2006) for the brain slice as given in Table 2 and displayed in Figure 7 were used. The units of measurement for T_2^* and T_1 are msec. In generating these data, a simplified version of (13) was used where the MFI $\Delta B(x, y)$ was not included. Without the inclusion of $\Delta B(x, y)$ MFI, the k-space array after being reconstructed yielded a real-valued image with a maximum around one. The average value in gray matter of this image was computed, and the entire image multiplied by a magnitude signal-to-noise ratio $\text{SNR}_M \times \sigma /\overline{\text{GM}}$, where $\sigma = 1$ is the standard deviation of noise added to the simulation, $\text{SNR}_M = 25$ is the signal-to-noise ratio for the simulation, and $\overline{\text{GM}} = 0.3545$ is the average gray matter value before scaling. This scaled real-valued image was used as the magnitude of the true images. To have nonzero mean phase, baseline phase as shown in Figure 8(b) was added to each tissue type according to $\alpha_{0,\text{OB}} = 0$ for outside brain, $\alpha_{0,\text{WM}} = \pi/12$ for white matter, $\alpha_{0,\text{GM}} = \pi/6$ for gray matter,

and $\alpha_{0,\text{CSF}} = \pi/4$ for cerebrospinal fluid. The remaining imaging parameters were selected to mirror those of an experimental dataset with field of view (FOV) = 240 mm, time to echo (TE) = 50 msec, flip angle (FA) = 90°, effective echo spacing (EESP) = 720 μsec, and bandwidth (BW) = 125 kHz.

The simulated data have slices of dimension 96×96 over $T = 490$ time points. The true activation regions are the two 5×5 red squares shown in the left panel of Figure 9. Each active voxel has different intensity and the voxels near the center of the region have stronger intensities than the ones around the edges of the region. At each time point, the magnitude contrast (β_1) in Figure 8(c) was multiplied by a task response waveform and then added to the magnitude baseline (β_0) in Figure 8(a) to form the image magnitude. At each time point, the phase contrast (α_1) in Figure 8(d) was multiplied by a task response waveform then added to the phase baseline (α_0) in Figure 8(b) to form the image phase. Independent zero mean and unit variance normal noise was also added to the real and imaginary parts at each time point. In this simulation, the maximum magnitude CNR in the center of each ROI was set to $\text{CNR}_M = \beta_1/\sigma = 0.5/1$ and the maximum phase CNR in each ROI was set to $\text{CNR}_P = \alpha_1/\text{SNR}_M = (\pi/120)/25$. The contrast values ($\beta_1$ and α_1) in each ROI were then multiplied by an unnormalized Gaussian kernel with full-width-at-half-max (FWHM) = 4 voxels.
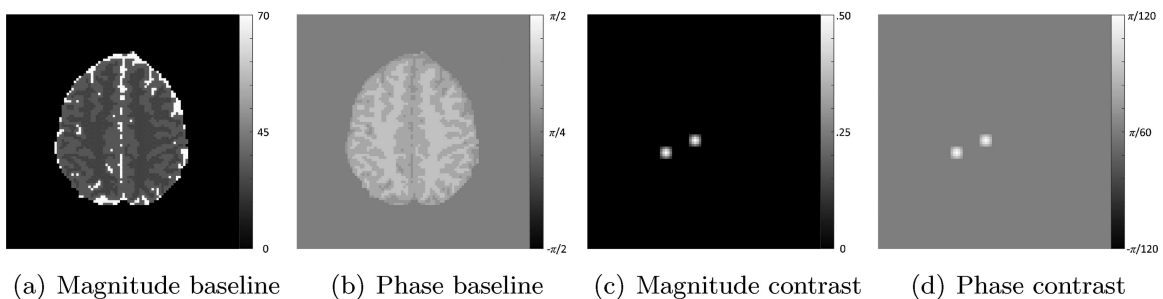
We fitted a Bayesian complex-valued (CV) model given by

$$y_t^v = \gamma_1^{*,v} + \gamma_2^{*,v}t/T + \gamma_3^{*,v}x_t + \eta_t^v, \quad \eta_t^v \sim \text{CN}_1\left(0, 2\sigma^2, 0\right), \tag{14}$$
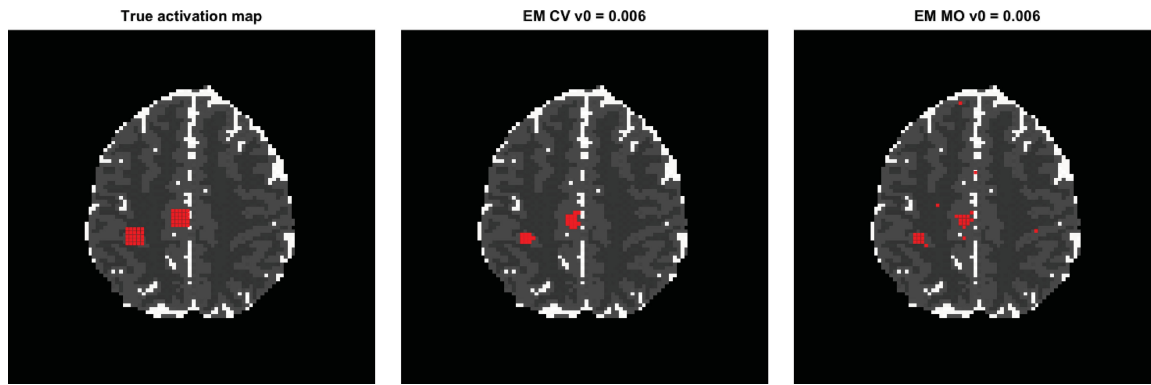
with $\gamma_1^{*,v}$ and $\gamma_2^{*,v}$ baseline and trend parameters, and $\gamma_3^{*,v}$ the activation parameters for $v = 1, \ldots, V$. Here, we set x_t to be the BOLD response obtained from the convolution of the experimental block design and the canonical HRF. We used priors of the form $\pi(\gamma_1^{*,v}) \propto 1$ and $\pi(\gamma_2^{*,v}) \propto 1$ for the baseline and trend parameters. For the activation parameters

(a) PSD, $\rho(x, y)$ (b) TRR, $T_2^*(x, y)$ (c) LRR, $T_1(x, y)$ (d) Time map, $t(x, y)$
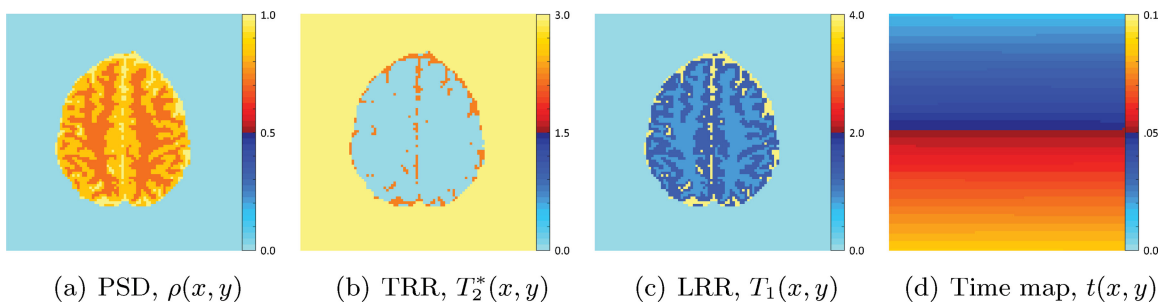
Figure 7. Slice physical parameters.

(a) Magnitude baseline   (b) Phase baseline   (c) Magnitude contrast   (d) Phase contrast

Figure 8. True simulated image values.

Figure 9. Left: True activation map. Middle: Activation map from C-EMVS at optimal v_0. Right: Activation map from magnitude-only EMVS at optimal v_0.

Table 3. Performance measures obtained from the complex-valued EMVS and magnitude-only EMVS in simulation study II.

Model	True positives (50 active)	False positives (9166 nonactive)	Precision	Accuracy	MSE (all)	MSE (active)
CV	24	0	1.000	0.9972	0.0046	0.0147
MO	23	5	0.821	0.9965	0.1582	0.0915

$\gamma_3^{*,v}$ for $v = 1, \ldots, V$, we used priors of the form $\gamma_3^{*,v} \mid \psi^v \sim (1 - \psi^v) CN_1(0, 2v_0\sigma^2, 0) + \psi^v CN_1(0, 2v_1\sigma^2, 0)$, and for σ^2, ψ^v, and θ we used the priors given in (12). As done in the simulation study I, we also fitted a magnitude-only version of this model (MO). The posterior results for both models summarized here were obtained after integrating out the baseline and trend parameters. In both models, we chose the optimal values of v_0 by maximizing the marginal posterior $\pi_0(\boldsymbol{\psi} \mid \mathbf{y})$ as described in the previous simulation study. In this case, the optimal v_0 values were found to be $v_0^{\mathrm{MO}} = v_0^{\mathrm{CV}} = 0.006$.

The posterior activation maps for each model are shown in Figure 9. First note that both Bayesian models, the complex-valued model (CV) and the magnitude-only one (MO), perform reasonably well in terms of detecting active sites, particularly considering that the CV-fMRI data were not generated from these models and instead followed a much more complicated physically realistic model, and also considering the low SNR and CNR in this setting. The main advantage of the Bayesian CV and MO models is that their linear structure allows us to obtain posterior estimates in a computationally feasible manner that scales well with the large dimension of the images.

Regarding the comparison between the complex-valued and magnitude-only models, we see that, once again, the CV model shows a better performance than the MO model. The MO model produces a larger number of false positives without detecting more true positives than the CV model. Table 3 shows the performance measures for both models. We also see that the mean squared errors are smaller for the CV model.

In terms of the strength, the CV model also leads to more accurate results. Figure 10 shows the estimated strengths obtained from the C-EMVS and MO EMVS approaches with their corresponding optimal v_0 values. The magnitude-only model overestimates the strengths for the true active sites and does not appropriately capture the fading effect observed in the true strength map. Finally, we note that full posterior results obtained via MCMC (not shown) were similar to those obtained through the EM approaches for both models.

4. Analysis of Human CV-fMRI Data

We analyzed human data recorded during an fMRI experiment performed on a single subject on a 3.0-T General Electric Signa LX magnetic resonance imager. The experiment consisted of a unilateral finger-tapping task performed with a visual cue indicating whether to tap or rest. A block designed experiment with an initial 20 sec of rest followed by 16 epochs of 15 sec on and 15 sec off was used. The full dataset is composed of seven 2.5 mm thick axial slices of dimension 96×96. A single slice was used for the analysis presented here. Further details about the experiment, the dataset and previous analyses of these data
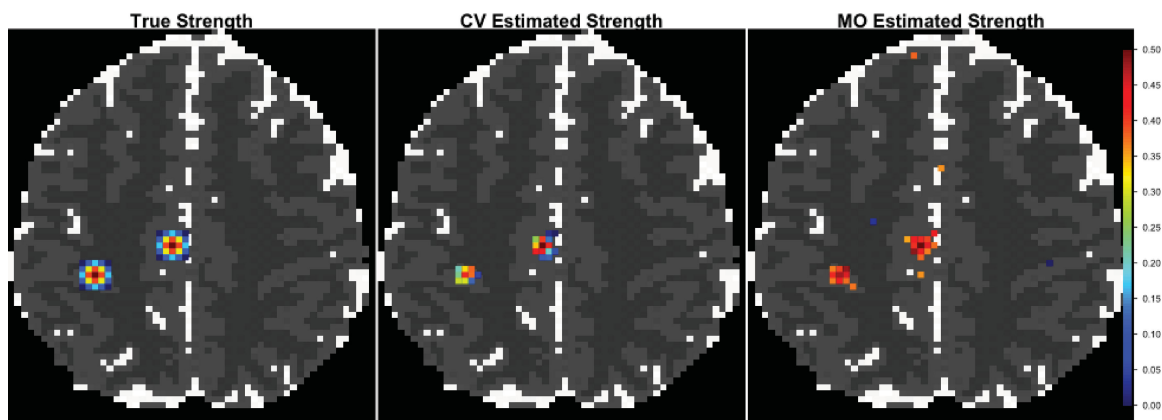
Figure 10. Left: True strength map. Middle: Strength map from C-EMVS at optimal v_0. Right: Strength map from magnitude-only EMVS at optimal v_0.

Table 4. CV models considered for the human CV-fMRI data.

Error structure	Common variance	Voxel-specific variance
independent	model (i): σ^2	model (iv): σ_v^2
AR(1), common AR parameter	model (ii): σ^2, φ	model (v): σ_v^2, φ
AR(1), voxel-specific AR parameter	model (iii): σ^2, φ_v	model (vi): σ_v^2, φ_v

are found in Karaman, Bruce, and Rowe (2014). The original time series at each voxel had 510 TRs, however, following the approach of Karaman, Bruce, and Rowe (2014), we discarded the first 20 observations for the analysis with the C-EMVS approach. Activation from this finger-tapping experiment is well-studied. However, the methods that have been used so far could have limitations in detecting activation—as suggested by the simulation studies. Our goal here is to demonstrate that our novel Bayesian complex-valued method is able to simultaneously produce activation results that are consistent with previous results and additionally lead to a reduction of spurious results, such as activation outside of the brain or in regions that are not implicated in the finger tapping task.

Karaman, Bruce, and Rowe (2014) analyzed these data with three different models: a complex-valued constant phase activation model that linearly describes the temporally varying magnitude (we refer to this model as KBR14-CV), a similar magnitude-only activation model (KBR14-MO), and a nonlinear model referred to as DeTeCT-ING that incorporates tissue and imaging parameters T_1 and T_2^* into physical magnetization equation to model magnetic resonance (MR) magnetization. More specifically, the DeTeCT-ING model considers the physical nonlinear signal equation to model MR magnetization, uses the first scans of the CV-fMRI data to estimate the parameters T_1 and T_2^*, and incorporates these GM (gray matter) parameter values to detect active voxels. Further details about these models and related activation maps obtained by Karaman, Bruce, and Rowe (2014) are included in the Appendices.

We applied the C-EMVS approach to these human CV-fMRI data using models with baseline, trend, and activation parameters and considered different noise structures. As in the previous examples, we used reference priors on the baseline and trend parameters and the proposed complex-valued spike-and-slab prior on activation parameters. We also used the canonical HRF to obtain the BOLD signals for all the voxels. Other classes of HRFs were considered, as explained in Section 3, resulting in similar activation results to those presented here for the canonical HRF. Regarding the noise structure, we considered

independent noise and noise with a temporal correlation modeled in terms of an autoregressive process of order one or AR(1). The model with AR(1) noise was specified as follows,

$$y_t^v = \gamma_1^{v,*} + \gamma_2^{v,*} t/T + \gamma_3^{v,*} x_t + \eta_t^v,$$

$$\eta_t^v = \varphi_v \eta_{t-1}^v + \zeta_t^v, \quad \zeta_t^v \overset{iid}{\sim} CN_1\left(0, 2\sigma_v^2, 0\right),$$

where φ_v is the AR(1) coefficient for voxel v. For this model, we considered a prior structure with $\pi(\gamma_1^{*,v}) \propto 1$, $\pi(\gamma_2^{*,v}) \propto 1$, $\gamma_3^{*,v} \mid \psi^v \sim (1 - \psi^v)CN_1(0, 2v_0\sigma_v^2, 0) + \psi^v CN_1(0, 2v_1\sigma_v^2, 0)$, $\sigma_v^2 \sim IG(1/2, 1/2)$, $\psi^v \mid \theta \sim Bernoulli(\theta)$, $\theta \sim Beta(1, 1)$, and $\varphi_v \sim U(-1, 1)$. In addition, we also considered models with common variance for all voxels, that is, $\sigma_v^2 = \sigma^2$ for all v and $\sigma^2 \sim IG(1/2, 1/2)$ and models with common AR coefficient for all voxels, that is, $\varphi_v = \varphi$ with $\varphi \sim U(-1, 1)$. All the different models that were considered are summarized in Table 4.

Here, we only present the results for models (iii) and (vi). We found that these two models led to better activation maps (i.e., smaller numbers of clear false positives in areas outside the brain) than the other models considered. The left and center plots in Figure 11 show the estimated values of φ_v for models (iii) and (vi). These pictures demonstrate that there is a large variability in the estimated AR coefficients at the voxel level. The voxels outside the brain essentially show no temporal correlation since the estimated AR coefficient values are close to zero. We also see that some voxels have relatively large temporal correlation with $\hat{\varphi}_v$ around 0.6, however these voxels do not lie in the activation areas. Figure 11 also shows the estimated values of σ_v^2 from model (vi) (right plot). It is clear from this plot that the estimated σ_v^2 are larger for those voxels inside the brain than for those outside. These estimated values are also able to differentiate gray matter from the rest and are consistent with results in Karaman, Bruce, and Rowe (2015). In particular, the right plot in Figure 11 shares similarities with the estimated T_1 map in Karaman, Bruce, and Rowe (2015). This is an important result given that our proposed models are able to capture a relatively sophisticated brain structure without incorporating nonlinear physically based components that would make posterior computations extremely challenging for these large dimensional data.

Figure 12 shows the strength maps obtained from models (iii) (left plot) and (vi) (right plot). These maps are fairly similar for both models. Comparing these C-EMVS results with those results obtained from the model of Karaman, Bruce, and Rowe (2014), we observe that the C-EMVS models have higher power of detecting active sites than the magnitude-only
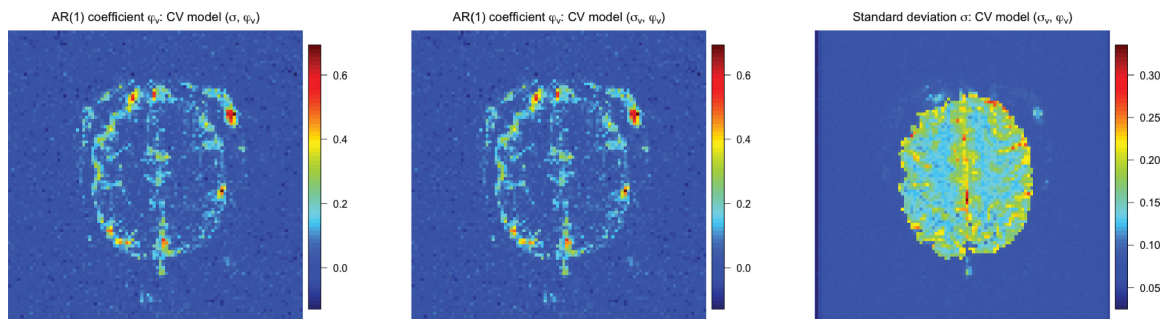
Figure 11. Human data: Estimated values of φ_v^2 for the CV model with a single σ^2 (model (iii), left plot) and the CV model with voxel-specific σ_v^2 (model (vi), center plot); estimated values of the φ_vs in model (vi).
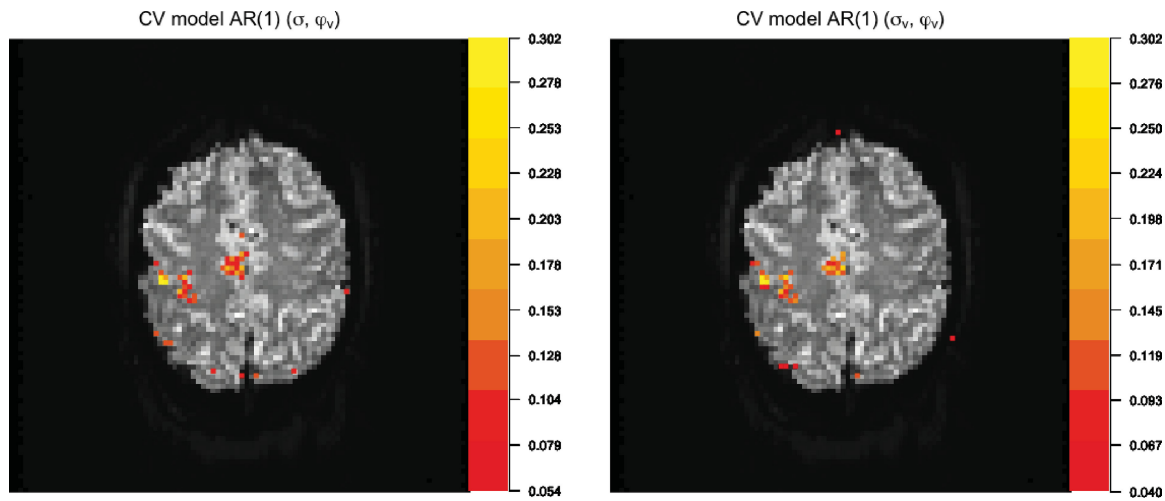
Figure 12. Human data: Strength maps obtained from models (iii) (left) and (vi) right.

KBR14-MO model, and also show a better performance than the complex-valued KBR14-CV model (KBR14 maps provided in the online Appendices). The maps obtained from models (iii) and (vi) show either no false positives outside of the brain and also no false positives in the upper left side of the brain close to the no signal area (model (iii)), or a much more reduced number of false positives (model (vi)) when compared to the activation map obtained from model KBR14-CV. In addition, The KBR14-MO and KBR14-CV models both use an FWE of 5% and are therefore procedures that require two steps. The Bayesian C-EMVS approach is a one-step procedure and does not require additional adjustments for multiple comparisons. The C-EMVS approach also compares favorably with the more physically realistic nonlinear DeTeCT-ING model. Models (iii) and (iv) identify most of the sites in the left and supplementary motor region and produce none or a reduced number of the false positive sites in the anterior left side of the brain than those that were obtained by the DeTeCT-ING model. These findings are not trivial especially given that, unlike the DeTeCT-ING model, the proposed Bayesian C-EMVS approach does not incorporate any physical aspects to model MR magnetization. This example shows that methods with improved power for detecting activations, such as the one developed here, are essential to increase the understanding of human brain function, particularly in scenarios that involve CV-fMRI images with low SNR.

5. Conclusion

Our main contribution in this article is a new Bayesian variable selection approach for detection of brain activation from single or multi-task complex-valued fMRI signals at the voxel-specific level. Although we focused on circular complex-valued priors and the methods were only illustrated in the context of CV-fMRI data, the models and algorithms proposed here are general, and can be applied to the general case of noncircular complex-valued priors and to other types of data.

Our simulation studies show that by considering both, real and imaginary information, the Bayesian complex-valued variable selection methods are able to detect more true positives and less false positives than magnitude-only models, especially when the SNR is small. We also found that both, the Bayesian complex-valued and magnitude-only EMVS approaches performed better than lasso and adaptive lasso and were computationally fast, with run times comparable to those needed by lasso or adaptive lasso. Finally, we demonstrated that the activation results in the finger-tapping experiment obtained from the C-EMVS approach compared favorably to those results obtained from more sophisticated nonlinear models that are physically realistic as they incorporate tissue and imaging parameters. The computational efficiency and the performance obtained in the analysis of experimental and simulated complex-valued fMRI data presented here make the C-EMVS approach a very useful tool for detecting brain activation.

We note that the new Bayesian complex-valued models considered here do not use any sophisticated spatio-temporal structure that can more appropriately describe fMRI data (we only considered an AR(1) temporal structure). Adding spatio-temporal structure that can better describe the data could potentially lead to further improved results, but would also lead to more computationally intensive models that may be not be feasible for detecting brain activation at the voxel-specific level. Future work will explore Bayesian complex-valued spatio-temporal extensions that are computationally scalable as well as multi-subject models. The C-EMVS methods presented here serve as a highly useful starting point, especially for analyzing high-dimensional CV-fMRI data.

Supplementary Materials

The online supplementary materials contain the appendices, and the associated R and C++ code for the C-EMVS algorithm. The simulated data used in Section 3.2 can be downloaded at *https://github.com/chenghanyu/CEMVS_simulation*

Funding

ORCID

Cheng-Han Yu http://orcid.org/0000-0002-6977-6400

References

Adrian, D. W., Maitra, R., and Rowe, D. B. (2013), "Ricean Over Gaussian Modelling in Magnitude fMRI Analysis–Added Complexity With Negligible Practical Benefits," *Stat*, 2, 303–316. [1396]

Adrian, D. W., Maitra, R., and Rowe, D. B. (2017), "Complex-Valued Time-Series Modeling for Improved Accuracy of Activation Detection in fMRI Studies," *Annals of Applied Statistics*, Accepted. [1396]

Bandettini, P. A., Wong, E. C., Hinks, R. S., Tikofsky, R. S., and Hyde, J. S. (1992), "Time Course EPI of Human Brain Function During Task Activation," *Magnetic Resonance in Medicine*, 25, 390–397. [1395]

Barbieri, M. M., and Berger, J. O. (2004), "Optimal Predictive Model Selection," *Annals of Statistics*, 32, 870–897. [1399]

Bernstein, M. A., Thomasson, D. M., and Perman, W. H. (1989), "Improved Detectability in Low Signal-to-Noise Ratio Magnetic Resonance Images by Means of a Phase-Corrected Real Reconstruction," *Medical Physics*, 16, 813–817. [1395]

Bowman, D. F., Caffo, B., Bassett, S. S., and Kilts, C. (2008), "A Bayesian Hierarchical Framework for Spatial Modeling of fMRI Data," *NeuroImage*, 39, 146–156. [1396]

Brandwood, D. H. (1983), "A Complex Gradient Operator and Its Application in Adaptive Array Theory," *Communications, Radar and Signal Processing, IEE Proceedings F*, 130, 11–16. [xxxx]

Calhoun, V. D., Adali, T., Pearlson, G. D., van Zijl, P. C. M., and Pekar, J. J. (2002), "Independent Component Analysis of fMRI Data in the Complex Domain," *Magnetic Resonance in Medicine*, 48, 180–192. [1396]

Chiang, S., Guindani, M., Yeh, H. J., Haneef, Z., Stern, J. M., and Vannucci, M. (2017), "A Hierarchical Bayesian Model for Identification of PET Markers Associated to the Prediction of Surgical Outcome After Anterior Temporal Lobe Resection," *Frontiers in Neuroscience*, 11, 1–16. [1396]

Friedman, J., Hastie, T., and Tibshirani, R. (2010), "Regularization Paths for Generalized Linear Models via Coordinate Descent," *Journal of Statistical Software*, 33, 1–22. [1401]

Friston, K. J., Ashburner, J. T., Kiebel, S. J., Nichols, T. E., and Penny, W. D. (2007), *Statistical Parametric Mapping: The Analysis of Functional Brain Images*, London: Academic Press. [1396]

Friston, K. J., Jezzard, P., and Turner, R. (1994), "Analysis of Functional MRI Time-Series," *Human Brain Mapping*, 1, 153–171. [1395]

George, E. I., and McCulloch, R. E. (1993), "Variable Selection via Gibbs Sampling," *Journal of the American Statistical Association*, 88, 881–889. [1397,1399]

—— (1997), "Approaches for Bayesian Variable Selection," *Statistica Sinica*, 7, 339–373. [1397]

Gudbjartsson, H., and Patz, S. (1995), "The Rician Distribution of Noisy MRI Data," *Magnetic Resonance in Medicine*, 34, 910–914. [1396]

Haacke, E., Brown, R., Thompson, M., and Venkatesan, R. (1999), *Magnetic Resonance Imaging: Principles and Sequence Design*, New York: Wiley. [1395,1405]

Hjørungnes, A., and Gesbert, D. (2007), "Complex-Valued Matrix Differentiation: Techniques and Key Results," *IEEE Transactions on Signal Processing*, 55, 2740–2746. [xxxx]

Karaman, M., Bruce, I. P., and Rowe, D. B. (2015), "Incorporating Relaxivities to More Accurately Reconstruct MR Images," *Magnetic Resonance Imaging*, 33, 374–384. [1404,1407]

Karaman, M. M., Bruce, I. P., and Rowe, D. B. (2014), "A Statistical fMRI Model for Differential T2* Contrast Incorporating T1 and T2 of Gray Matter," *Magnetic Resonance Imaging*, 32, 9–27. [1407]

Kociuba, M. C., and Rowe, D. B. (2016), "Complex-Valued Time-Series Correlation Increases Sensitivity in fMRI Analysis," *Magnetic Resonance Imaging*, 34, 765–770. [1396]

Lai, S., and Glover, G. H. (1997), "Detection of BOLD fMRI Signals Using Complex Data," in Proceedings of the 5th Annual Meeting of ISMRM, Vancouver, BC, Canada, p. 1671. [1396]

Lee, J., Shahram, M., and Pauly, J. M. (2009), "Combining Complex Signal Change in Functional MRI," *Magnetic Resonance in Medicine*, 62, 1358–1360. [1396]

Lee, J., Shahram, M., Schwartzman, A., and Pauly, J. M. (2007), "Complex Data Analysis in High-Resolution SSFP fMRI," *Magnetic Resonance in Medicine*, 57, 905–917. [1396,1397]

Lee, K. J., Jones, G. L., Caffo, B. S., and Bassett, S. S. (2014), "Spatial Bayesian Variable Selection Models on Functional Magnetic Resonance Imaging Time-Series Data," *Bayesian Analysis*, 9, 699–732. [1396]

Macovski, A. (1996), "Noise in MRI," *Magnetic Resonance in Medicine*, 36, 494–497. [1395]

Mandic, P. D., and Goh, V. S. L. (2009), *Complex-valued Nonlinear Adaptive Filters*, New York: Wiley. [1396]

Müeller, P., Parmigiani, G., and Rice, K. (2006), "FDR and Bayesian Multiple Comparisons Rule," in *Bayesian Statistics 8: Proceedings of the Eighth Valencia International Meeting*, eds. J. M. Bernardo, M. J. Bayarri, J. O. Berger, A. P. Dawid, D. Heckerman, A. F. M. Smith, and M. West, Oxford: Oxford University Press. [1399]

Oppenheim, A. V., and Lim, J. S. (1981), "The Importance of Phase in Signals," *Proceedings of the IEEE*, 69, 529–541. [1396]

Peters, A. M., Broooks, M. J., Hoogenrad, F. G., Gowland, P. A., Francis, S. T., Morris, P. G., and Bowtell, R. W. (2006), "Comparison of T2* Measurements in Human Brain at 1.5, 3 and 7 T," in *Proceedings of the International Society for Magnetic Resonance in Medicine* (Vol. 14), Seattle, WA: Wiley, p. 926. [1405]

Picinbono, B. (1996), "Second-Order Complex Random Vectors and Normal Distributions," *IEEE Transactions on Signal Processing*, 44, 2637–2640. [1397]

Prah, D. E., Paulson, E. S., Nencka, A. S., and Schmainda, K. M. (2010), "A Simple Method for Rectified Noise Floor Suppression: Phase-Corrected Real Data Reconstruction With Application to Diffusion-Weighted Imaging," *Magnetic Resonance in Medicine*, 64, 418–429. [1395]

Rice, S. O. (1944), "Mathematical Analysis of Random Noise," *The Bell System Technical Journal*, 23, 282–332. [1396]

Rockova, V., and George, E. I. (2014), "EMVS: The EM Approach to Bayesian Variable Selection," *Journal of the American Statistical Association*, 109, 828–846. [1397,1398,1399,1401]

Rowe, D. B. (2005a), "Modeling Both the Magnitude and Phase of Complex-Valued fMRI Data," *NeuroImage*, 25, 1310–1324. [1396,1397]

—— (2005b), "Parameter Estimation in the Magnitude-Only and Complex-Valued fMRI Data Models," *NeuroImage*, 25, 1124–1132. [1396,1397,1402]

—— (2009), "Magnitude and Phase Signal Detection in Complex-Valued fMRI Data," *Magnetic Resonance in Medicine*, 62, 1356–1357. [1396,1397]

Rowe, D. B., and Logan, B. R. (2004), "A Complex Way to Compute fMRI Activation," *NeuroImage*, 23, 1078–1092. [1396,1397,1400]

—— (2005), "Complex fMRI Analysis With Unrestricted Phase is Equivalent to a Magnitude-Only Model," *NeuroImage*, 24, 603–606. [1396,1397]

Scott, J. G., and Berger, J. O. (2006), "An Exploration of Aspects of Bayesian Multiple Testing," *Journal of Statistical Planning and Inference*, 136, 2144–2162. [1397]

Smith, M., and Fahrmeir, L. (2007), "Spatial Bayesian Variable Selection With Application to Functional Magnetic Resonance Imaging," *Journal of the American Statistical Association*, 102, 417–431. [1399]

Sun, W., Reich, B., Cai, T., Guindani, M., and Schwartzman, A. (2015), "False Discovery Control in Large-Scale Spatial Multiple Testing," *Journal of the Royal Statistical Association*, Series B, 77, 59–83. [1399]

Tibshirani, R. (1996), "Regression Shrinkage and Selection via the Lasso," *Journal of the Royal Statistical Society*, Series B, 58, 267–288. [1399,1401]

van den Bos, A. (1995), "The Multivariate Complex Normal Distribution—A Generalization," *IEEE Transactions on Information Theory*, 41, 537–539. [1397]

Wang, J., Liang, F., Li, Y., and Zhu, Y. (2015), "A Scalable Algorithm for Bayesian Variable Selection (SAR) with Application to miRNA-mRNA Regulation in Cancer," submitted. [1398,1401]

Welvaert, M., Durnez, J., Moerkerke, B., Verdoolaege, G., and Rosseel, Y. (2011), "neuRosim: An R Package for Generating fMRI Data," *Journal of Statistical Software*, 44, 1–18. [1399]

Wooding, R. A. (1956), "The Multivariate Distribution of Complex Normal Variables," *Biometrika*, 43, 212–215. [1397]

Xia, J., Liang, F., and Wang, Y. (2009), "FMRI Analysis Through Bayesian Variable Selection With a Spatial Prior," in *Proceedings of the 2009 IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, Boston, MA: IEEE, pp. 714–717. [1397]

Yu, Z., Prado, R., Ombao, H., Quinlan, E., Cramer, S., and Ombao, H. (2016), "Understanding the Impact of Stroke on Brain Motor Function: A Hierarchical Bayesian Approach," *Journal of the American Statistical Association*, 111, 549–563. [1403]

Zhang, L., Guindani, M., and Vannucci, M. (2015), "Bayesian Models for fMRI Data Analysis," *WIREs Computational Statistics*, 7, 21–41. [1396,1397]

Zhang, L., Guindani, M., Versace, F., Engelmann, J. M., and Vannucci, M. (2016), "A Spatio-Temporal Non-Parametric Bayesian Model of Multi-Subject fMRI Data," *Annals of Applied Statistics*, 10, 638–666. [1396]

Zhang, L., Guindani, M., Versace, F., and Vannucci, M. (2014), "A Spatio-Temporal Nonparametric Bayesian Variable Selection Model of fMRI Data for Clustering Correlated Time Courses," *NeuroImage*, 95, 162–175. [1396]

Zou, H. (2006), "The Adaptive Lasso and Its Oracle Properties," *Journal of the American Statistical Association*, 101, 1418–1429. [1399,1401]